DOCUMENT RESUME

ED 062 901 24 FL 003 359

AUTHOR Wipf, Joseph Arnold
TITLE An Investigation of the Effect of Maturation on Imitative Ability in Second Language Learning: A Psycho-Linguistic Study. An Investigation of Elementary School Students' Ability to Imitate Selected Sound Features of German.
INSTITUTION Ohio State Univ., Columbus. Research Foundation.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Regional Research Program.
BUREAU NO BR-1-E-022
PUB DATE Apr 72
GRANT OEG-5-71-0026-509
NOTE 165p.; Final report, shortened version of a Ph.D. dissertation

EDRS PRICE MF-$0.65 HC-$6.58
DESCRIPTORS *Age Differences; Applied Linguistics; Educational Experiments; *Elementary Grades; Fles; German; *Imitation; Language Instruction; Language Laboratories; *Language Research; Language Skills; Maturation; Phonology; Psycholinguistics; *Second Language Learning; Speech; Tables (Data)

ABSTRACT

In order to gain some insight into the issue of optimum age for second language learning, the experiment described in this report was based on the negative hypothesis that there is no significant difference in the ability of randomly selected subjects, kindergarten through grade six, to imitate critical sounds and sound clusters embedded in words and phrases, word stress, and sentence stress of German. This document provides background information on the issue and reviews related literature; experiment procedures are described and the statistical results, summary, and conclusions are presented. According to the findings, imitative ability increases with grade level, from kindergarten to grade 6. (VM)

FILMED FROM BEST AVAILABLE COPY

ED 062 901

Final Report

Project No. 1-E-022
Grant No. OEG-5-71-0026(509)

AN INVESTIGATION OF THE EFFECT OF MATURATION ON IMITATIVE ABILITY IN SECOND LANGUAGE LEARNING: A PSYCHO-LINGUISTIC STUDY

(An Investigation of Elementary School Students' Ability to Imitate Selected Sound Features of German)

Joseph Arnold Wipf

The Ohio State University
Research Foundation

Columbus, Ohio

April 1972

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Regional Research Program

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FL 003 359

AN INVESTIGATION OF THE EFFECT OF MATURATION ON IMITATIVE ABILITY IN SECOND LANGUAGE LEARNING: A PSYCHO-LINGUISTIC STUDY

(An Investigation of Elementary School Students' Ability to Imitate Selected Sound Features of German)

(Shortened Version of Ph.D. Dissertation)

The purpose of this study was to identify the relationship between subjects from various grade levels and their ability to imitate selected sound features of German. The subjects consisted of two hundred ten randomly selected children, fifteen boys and fifteen girls from each grade level, kindergarten through grade six. These subjects were asked to imitate forty German utterances arrived at on the basis of a pilot test conducted with thirty-nine subjects of the same age range. The test stimuli consisted of twenty monosyllabic words and twenty phrases ranging in length from two to four syllables. Only one critical sound or sound cluster was embedded in each of the utterances which were modeled for prerecording by a native German linguist.

The test was administered in two of the new and ultra-modern language laboratories located in the Listening Center in the Dieter Cunz Hall of Languages on The Ohio State University Campus. Test stimuli were modeled three times. Responses of the subjects were recorded on magnetic tape and were subsequently scored by two highly qualified native-speaking judges. The production of individual sounds and sound clusters was scored on a six-point scale and word and sentence stress were evaluated on a three-point scale.

An analysis of three variables (judges, sex and grade level) was made by subjecting the test scores to two statistical tests: Mann-Whitney U and Fisher-Student t. Based on scores achieved on the total instrument designed for this study, the following conclusions were reached:

1. There was no significant difference in the scores assigned to the subjects by the two judges.

2. There was no significant difference in the scores achieved by the two sexes.

3. There was a significant difference in the ability of the subjects from the seven grade levels to imitate selected sound features of German. This statement is made on the grounds that:

   a. The sixth graders ranked highest with scores significantly better than all other grade levels.

   b. The third, fourth and fifth graders did equally well and ranked second having outscored all three lower grades at a significant level.

c. The first and second graders rated third. First graders were slightly better since they achieved a significantly better score than the kindergarteners.

d. The kindergarteners were the weakest subjects having a significantly lower score than subjects from all other grades except the second.

A comparison of the top ten subjects from each level produced similar results. There was no significant difference in judge variability, nor in achievement according to sex. Although this analysis did not produce quite as marked differences among grade levels, the same general pattern appeared--imitative ability increased with grade level.

This writer wishes to emphasize that the ability to imitate critical sound features of a foreign language is just one aspect (but an important one) of second language acquisition. Other factors such as the mastery of vocabulary, morphology and syntax must not be overlooked. Nor was it the intent of this investigation to measure the ability of subjects from various grade levels to learn how to imitate the sound features of German over a period of time in an instructional situation. Further research is also necessary to determine what accounts for the significant differences among grade levels in this study. For example, the imitative task might depend on factors such as: acoustic perception, coding and memory in addition to actual articulatory functioning. To make sweeping generalizations on the basis of one study would be a mistake of serious dimensions--one which a profession treading on delicate soil can ill afford.

TABLE OF CONTENTS

		Page
ACKNOWLEDGMENTS		ii
VITA		iv
LIST OF TABLES AND GRAPHS		viii
Chapter		
I	INTRODUCTION	1
	Rationale	1
	The Problem	4
	Hypothesis	7
	Definition of Terms	8
	Description of Phonetic Transcription & Diacritic Marks	10
II	REVIEW OF RELATED LITERATURE	12
III	PROCEDURES	27
	Population and Sample	27
	Instrumentation	28
	Data Analysis	50
	Personnel	52
IV	RESULTS	55
	Pilot Test	55
	Final Test	63
	Variable 1: Judges	63
	Variable 2: Grade Level	68
	Part 1A: Sounds and Sound Clusters in Single Words	68
	Part 1B: Word Stress	75
	Part 2A: Sounds and Sound Clusters in Phrases	76
	Part 2B: Sentence Stress	82
	Total Test	83
	Reliability	87

TABLE OF CONTENTS (Cont'd)

		Page
	Variable 3: Sex	87
	Comparison of the Ten Best Subjects On Each Grade Level	87
	Ranking of Sounds and Sound Clusters by Difficulty	91
V	SUMMARY AND CONCLUSIONS	95
	Generalizability	99
	Implications for Instructional Strategies and Materials	101
	Implications for Further Research	104
APPENDIXES		106
A	Score Sheet: Pilot Sound Production Test	106
B	Directions for Pilot Test	108
C	Letter to Subjects for Pilot Test	111
D	Response Sheet from Subjects of Pilot Test	114
E	Score Sheet: Final Sound Production Test	116
F	Directions for Final Test	118
G	Vita of Personnel	121
H	Explanation of Computer Print-out Labels	134
I	Data Analysis (Computer Print-outs): All Subjects	136
J	Data Analysis (Computer Print-outs): Top 10 Subjects from Each Grade Level	145
BIBLIOGRAPHY		154

LIST OF TABLES AND GRAPHS

Table		Page
1	Pilot Test: Correlation of Judge Agreement on Part 1A	60
2	Pilot Test: Correlation of Judge Agreement on Part 1B	62
3	Final Test: Judge Variability on Parts 1A and 1B	64
4	Part 1A: Sounds Embedded in Single Words--Significant Differences Among Grade Levels on Each Item	69
5	Part 1A: Cross-Tabulation of Significant Differences Among Grade Levels on Individual Items	70
6	Part 1A: Significant Differences Among Grade Levels on Composite Scores	72
7	Part 1B: Word Stress: Significant Differences Among Grade Levels on Composite Scores	76
8	Part 2A: Sounds Embedded in Phrases: Significant Differences Among Grade Levels on each Item	77
9	Part 2A: Cross-Tabulation of Significant Differences Among Grade Levels on Individual Items	78
10	Part 2A: Significant Differences Among Grade Levels on Composite Scores	79
11	Part 2B: Sentence Stress: Significant Differences Among Grade Levels on Composite Scores	83
12	Total Test: Significant Differences Among Grade Levels	84

Table		Page
13	Rank, Means and Standard Deviations by Grade Level on Total Test	85
14	Top Ten Subjects on Each Grade Level: Significant Differences Among Grade Levels	89
15	Mean Scores of the Top Ten Subjects in Each Grade Level	90
16	Rank Order of Sounds According to Difficulty	92
Graph		
1	Mean Score by Grade Level on Parts 1A and 2A	73
2	Mean Scores on Total Test	86

CHAPTER I

INTRODUCTION

Rationale

There is a great deal of controversy on the American educational scene today as to when a child should begin learning a second language in our schools. School boards, administrators and teachers have long been relying on subjective judgment as well as inconclusive and conflicting research in determining the appropriate time for beginning this activity. Consequently, the starting time for second-language learning varies from kindergarten to college.

With the advent of the audio-lingual era came an increased emphasis on learning to speak the foreign languages taught in our classrooms. As a result, skill in imitating foreign language sounds became an extremely significant index for measuring student success in foreign language learning. Modern curriculum innovations are stressing individualization of instruction (the main theme of Volume II of the Britannica Review of Foreign Language Education, Lange, 1970, p. 2), giving students the option to elect different types and concentrations of foreign language instruction, depending on the degree of proficiency desired as well as on the students' talents and interests (Arendt, 1970, p. 12).

Nevertheless, the fact remains, that speaking is usually a skill of high priority. This point of view has repeatedly been expressed by teachers and students alike. Werner Haas of The Ohio State University German Department has said recently (1970, p. 64):

> Advocating the emphasis of only certain skills can hardly be considered a progressive step in foreign language teaching. Such a stance could lead us back to the "reading only" phase of past decades, a phase which our profession has tried so hard to overcome.

Similarly, Wilga Rivers (1968, p. 220) advocates the teaching of the four skills concurrently, with greatest emphasis on practice in listening and speaking in the early stages. This, she maintains, provides for a greater variety of classroom activity for both teacher and student. Robert Lado in his book Language Testing (1964, p. 239) stresses the importance of the speaking skill even more emphatically.

> The ability to speak a foreign language is without a doubt the most highly prized language skill, and rightly so, because he who can speak a language well can also understand it and can learn to read it with relative ease . . . . Also, the ability to speak a language will greatly expedite and facilitate learning to write it.

Student subjects in Pimsleur's study on under-achievement (1966, p. 11) frequently said in interviews that they would like more class time devoted to speaking the foreign language, that speaking is fun, that it is important, and that they were motivated more by oral than by written work.

The importance of imitating foreign language sounds correctly when practicing the speaking skill was perhaps best summarized by the noted researcher John Carroll when he said:

> Children who can imitate foreign language sentences quickly and accurately are most likely to succeed. (May, 1960, p. 14.)
>
> For a person who hopes to communicate with the speakers of a foreign language, a reasonably good approximation to the accepted pronunciation of the language is necessary, and for the person who hopes to be taken as a "near-native" in fluency, a close approximation is absolutely essential. (1963, p. 1069.)

The urgency of displaying satisfactory skill in imitating new sounds <u>early</u> in the second-language experience must also not be overlooked. Agard and Dunkel (1948, p. 288) report that students entering an audio-lingual course after previous study of the language in high schools where English habits were tolerated, generally never succeeded in matching the pronunciation of those who were taught from the beginning to imitate native models.

And even if only reading and writing are the goals of the learner, it is felt by some that a surer mastery of these skills can be obtained if he passes through a substantial stage of work with the spoken language (Carroll, 1963, p. 1063).

The review of related literature will also show, that there is a real need for further research to determine the most appropriate time for the learner to begin successful second language study which often depends heavily on imitative ability in our classrooms today.

The Problem

In the United States, the late 50's and early 60's witnessed the expansion of our foreign language programs and curriculum into the elementary school. Enrollments increased throughout the latter decade to the point where FLES (Foreign Language in Elementary Schools) was offered by approximately ninety-five per cent of the large public schools systems (with 100,000 students or more), seventy-five per cent of the average systems (with 50,000 to 99,999 students), sixty per cent of the low average systems (with 25,000 to 49,999 students), and by 50 per cent of the small systems (with 12,000 to 24,999 students) reporting to the NEA Research Bureau in December 1967 (Donoghue, 1969, p. 1).

This movement is at least partially based on the belief that the ability to learn a foreign language declines with age (Grittner, 1969, p. 63). For this reason a great many FLES programs have begun in grades three or four when the children are eight or nine years of age. Some reasons for not starting earlier are that the child should first become well adjusted to the school world and that he needs a firm foundation in his mother tongue (Brooks, 1964, p. 117); furthermore the expense makes it prohibitive and there is a lack of competent teachers (JeKenta and Fearing, 1968, p. 142).

But it is rarely argued that the child's imitative ability at an earlier age is inferior. Examples of those who assert adult superiority in language acquisition are Ausubel (1967, p. 100) and

Kaulfers (Denemark and Matson, 1960). Yet it may well be the case that the capacity to imitate increases for a time with chronological age (Grinder, Otoma and Toyota, 1962). At any rate, if it could be established that there are clear-cut turning points in imitative ability in relation to chronological age, school officials and the foreign language teaching profession would be in a stronger position to recommend initial foreign language training at a given age level. With foreign language instruction presently being the target of severe criticism to the point where its very existence in the curriculum even on the college level is being challenged, a sound theoretical base on which to stand would be highly desirable. In addition, instructional strategies and materials at various age levels might be grossly affected.

There are, of course, other reasons advanced for teaching foreign languages at an early age. One of the most logical ones is that the earlier the start, the more years of formal training the child can get (Carroll, May, 1960, p. 13). A counter argument that has been raised, however, is that the degree to which children can be placed in advanced classes in high school foreign language classes because of previous experience in elementary school foreign languages is disappointing. The amount of advanced placement is seldom more than one year, while the amount of FLES experience has sometimes been as much as six years (Denemark and Matson, 1960).

Still other conjectures that have appeared in FLES literature are: a child has more time than adults, and he is not under so many distracting pressures (Grittner, 1969, p. 62); he is less inhibited in group learning and is more fascinated than adults by novel means of communication (Stern, 1966, pp. 253-264).

The following assertions which are used in support of FLES are, however, more intimately connected with the central issue of imitative ability: children are more receptive to new sounds because they don't immediately compare to English what they hear in the new language (Grittner, p. 63), and the child articulates new sound features better because his speech organs are still relatively flexible (Grittner, 1969, p. 63; Stern, 1966, p. 263). To the writer's knowledge, these hypotheses have never really been subjected to scientific investigation. Even if proof were demonstrated, it would not necessarily follow that maturity is in fact a severe handicap in imitative ability.

The problem then, which was investigated is: the relationship between grade level of students and their ability to imitate foreign language sounds. This was measured by asking randomly selected subjects ages five to eleven years, to imitate German utterances arrived at on the basis of a pilot test. Test stimuli included German sounds and sound clusters (embedded in words and phrases) that can cause the speaker of English difficulty. The responses of the subjects were recorded on magnetic tape and were subsequently evaluated by two native-speaking judges. The

scores were then analyzed to determine if the differences among the subject age groups were significant.

The writer wishes to substantiate the absence of and necessity for such research by means of the following quotations:

> It must be pointed out that the current views on the optimum age were not founded on systematic observation of children learning foreign languages under classroom conditions, but were extrapolations based on general knowledge of brain neurology and child development . . . . The conclusion we reach is that the current claim that the early years of schooling offer optimal conditions for language learning is open to question. The way it is commonly formulated implies that later learning in adolescent and adult life is not so good. Yet there is no evidence for this. (Stern, 1966, pp. 264-265.)

> There is no evidence to justify California's legislation for foreign languages in grades six to eight. (Cronbach, 1966, p. 541.)

> If research does not clearly indicate whether learning in general is better in childhood or adulthood, still less does it indicate exactly when FL instruction "should" begin. (Carroll, May, 1960, p. 13.)

> . . . . the question of the placement of foreign language instruction and the most appropriate years for beginning such instruction is not a settled matter but deserving of careful study and research. (Denemark and Matson, 1960, p. 11.)

Hypothesis

This study attempted to identify the relationship between subjects from various grade levels and their ability to imitate selected sound features of German. The following specific null hypothesis was tested: there is no significant difference in the

ability of randomly selected subjects, kindergarten through grade six, to imitate critical sounds and sound clusters embedded in words and phrases, word stress, and sentence stress of German. Subjects were of a similar socio-economic background, from the same geographic area and had no previous experience with or exposure to any foreign language other than what most children hear on popular children's television programs such as "Mister-rodgers' Neighborhood," "Sesame Street" and "Lucy's Toy Shop."

Definition of Terms

1) Audio-lingual: a term used to refer to listening and speaking and to teaching designed to produce these skills (Walsh, 1964, p. 12). All initial language learning and teaching occur via the skills of listening and speaking. The skills of reading and writing are presented later in the sequence.

2) Compound bilingual: a speaker whose target language is not sufficiently mastered to permit it to function as a system of communication independent of this person's native language (Brooks, 1964, p. 267).

3) Coordinate bilingual: a speaker who uses either his native language system or that of the target language, but not both simultaneously. The systems operate independently with only minor or incidental effects upon each other (Brooks, 1964, p. 267).

4) Critical sound: a non-English sound.

5) Elementary school: kindergarten through grade 6.

6) Imitative ability: the facility to reproduce accurately the segmental and supra-segmental features of the target language after having heard them modelled. In practice this would mean producing individual sounds and sound clusters correctly in terms of the correct manner and place of articulation and with the proper stress and intonation patterns.

7) Manner of articulation: the type of sound-producing or sound-modifying mechanisms in the mouth, such as plosives or fricatives (Gleason, 1967, p. 240).

8) Organ of articulation: a vocal organ such as the tongue or lower lip which is used to partially or wholly obstruct or to change the size and shape of the resonance chamber (Gleason, 1967, p. 242; Walsh, 1964, p. 11).

9) Point of articulation: place of maximum constriction in the mouth or pharynx, such as the teeth or lips (Gleason, 1967, p. 240).

10) Segmental features: vowels and consonants (Walsh, 1964, p. 29).

11) Sound cluster: two or more consonants or vowels in sequence (Walsh, 1964, p. 13).

12) Suprasegmental features: pitch and stress (Walsh, 1964, p. 30).

13) Target language: the language being studied (Walsh, 1964, p. 31).

Description of Phonetic Transcription and Diacritic Marks

[i:] = long, close i- sound.

[e:] = long, close e- sound.

[o:] = long, close o- sound.

[ɔ] = short, open o- sound.

[a] = short, front a- sound.

[ø:] = long, close o- umlaut

[œ] = short, open o- umlaut.

[y:] = long, close u - umlaut.

[Y] = short, open u- umlaut.

[a^{U}] = diphthong: short front a- sound plus a short open u- sound.

[ç] = voiceless, palatal fricative.

[x] = voiceless, velar fricative.

[l] = voiced, alveolar lateral.

[R] = voiced, uvular trill.

[pf] = affricate: a homeogeneous combination of a voiceless bilabial stop component and voiceless labio-dental fricative component.

[kn] = consonant cluster containing a voiceless, velar stop and a voiced, alveolar nasal.

[ʃt] = consonant cluster containing a voiceless prepalatal fricative and a voiceless alveolar stop.

[tsv] = three-element consonant cluster containing a voiceless alveolar stop, a voiceless alveolar fricative and a voiced labio-dental fricative.

[çts] = three-element consonant cluster containing a voiceless platal fricative, a voiceless, alveolar stop and a voiceless alveolar fricative.

[ ] = phonetic bracket.

: = full length as in the vowel [o:].

̣ = reduction of the sound as [ʀ̣] in "ihr."

CHAPTER II

REVIEW OF RELATED LITERATURE

Attention will first be drawn to relevant research done outside the domain of foreign language teaching per se, but which has implications for this study.

A large scale study (Roe and Milisen, 1942) of defective articulation of native English by elementary school children came to the conclusions that there is a significant reduction in the average number of errors made between grades one and two, two and three, and three and four. However, between grades four and five and again between grades five and six, they found no evidence of a continued reduction in errors. This would suggest that maturation or learning may not result in noticeable improvement in the speech sounds produced by pupils in the higher grades. Another researcher (Sayler, 1949), then extended the study to embrace grades seven through twelve. Only a slight amount of improvement was found in grades seven through ten, and essentially no improvement was detected in grades eleven and twelve. Although one cannot be certain, it seems likely that these patterns of improvement (as well as the levelling off of improvement) would have a relationship to second-language learning at the

same age levels. One could no doubt expect parallel patterns in children's ability to produce sounds of a second language. Wängler (in correspondence of February, 1971), has pointed out, however, that this would not necessarily be true in all languages, that it would depend on the sound systems involved, their articulatory complexity and difficulty and differences from English or the native language.

A study which was directed towards a comparison of the ability of children with good and poor articulation to produce sounds not present in the English language sheds some light on this problem (Winitz and Lawrence, 1961, pp. 259-268). The subjects of this study were selected from a group of ninety-six kindergarten children on the basis of their scores on the Templin Articulation and Screening Test. In order to experiment with two extreme groups in articulatory ability, the twelve subjects in the upper 12.5 per cent were identified as having good articulation (high group), and the twelve subjects in the lower 12.5 per cent as having poor articulation (low group). The task of the subjects was to produce three non-English sounds: [x], [œ] and [ç], after having heard them uttered numerous times by a native German linguist. The results of this study appeared to indicate that kindergarten children with good and with poor articulation are equally facile in learning to perform a sound task consisting of sounds not present in the English language. The explanation offered for these results is that differences in articulatory ability

may be due to some rather complex reinforcement contingencies of the past or present. When conditions of learning are similar, as they were in this study, differences between children with good and poor articulation are not apparent in the rate or level of learning. This research is of special interest here since the three non-English sounds tested in this instrument were German, all three of which appear in the instrument designed for the research conducted here. Furthermore, preliminary experimentation with children not included in the Winitz and Lawrence study demonstrated that kindergarten children could produce these sounds.

While really only a by-product of his research in neurology and neurosurgery, the publications of Dr. Wilder Penfield, the noted director of the Montreal Neurological Institute, are drawing more and more attention from the foreign language teaching profession (Miel, 1954, p. 143; Boehm, 1959, p. 32; Hildreth, 1959, p. 138; Larew, 1961, pp. 203-204; Michel, 1967, pp. 192-214; Grittner, 1969, pp. 63-64). Penfield's research (1953, pp. 199-214) has uncovered evidence that there is an age when the child has a remarkable capacity to utilize the four separate areas of the human cerebral cortex for the learning of language. During this time several languages can be learned simultaneously as easily as one language. When the capacity for reason and abstract thinking appears, this early ability is lost. Similarly in a more recent publication (1964, pp. 77-81),

Penfield explains an adult's failure (as compared to some children's success) in recovering full control of his speech after suffering brain damage. This is apparently due to the fact that by adulthood a person has taken over the initially uncommitted convolutions of his brain for other uses. Penfield (1959, p. 255), recommends the ages between four and ten as the time to begin what he calls a general schooling in a secondary language.

Grittner (1969, p. 64), suggests that the implications of the neurological evidence are that only a child of ten or younger can ordinarily become a full coordinate (not compound) bilingual, because someone who reaches adulthood as a monolingual is compelled to superimpose the second language upon areas of the human brain already committed to the learner's native language. He may become a compound bilingual commanding a considerable vocabulary, but his accent will betray him as someone who commenced his language study after the optimum age.

It is interesting to note the reaction of a number of other leading neurologists and psychiatrists to Penfield's statement made in 1953 (that language is learned more easily by young children). Ten specialists in the field of neurology and three in psychiatry were asked whether they agreed with this point of view. Seven of the neurologists supported Penfield's assertion (some with qualifications), while three of the

neurologists and all three of the psychiatrists took major exception to his statement (Foreign Language Bulletins, No. 1, 1953; Miel, 1954, pp. 143-144).

Thus it is evident that even outside the field of foreign language instruction there is anything but agreement on this controversial subject.

The research done in foreign languages proper also indicates that there is a definite difference of opinion as to the ideal time for commencing foreign language study. Since it is easy to confuse opinion with research-validated facts, only those publications will be discussed here which are based on concrete research findings, or at least on actual teaching experience with various age groups.

The advocates of an early start will be cited first. Kirch (1955, pp. 144-145; 1956, pp. 399-400) found an inverse relationship between elementary school age and ability to reproduce the sounds of German. Accordingly, he recommended that FLES instruction begin as early as possible--certainly by first grade. He bases his evidence on his experiences teaching German on three different grade levels: first grade children of a public school in Philadelphia, a third grade class in Newark, Delaware, and a sixth grade class, also in the latter school. The three different classes were taught in three successive years. Kirch maintains that the eleven-year olds had much more difficulty with the ch-sounds, the umlauted vowels and the German r-sound than the

younger ones did. He added, however, that most of them (eleven-year olds) finally produced these sounds correctly, but it took considerable time and effort. The first-graders were able to produce them properly after hearing them only two or three times (1956, p. 399).

Critics of Kirch's assertions (Grinder, Otoma and Toyota, 1962, p. 194) make the accusation that his results were obtained under conditions that were likely to have facilitated a superior performance by the first graders. Some of them were motivated to learn German because they wanted to communicate with their German-speaking parents and grandparents. They no doubt learned some pronunciation skills at home. In addition the first grade curriculum stressed concepts about the immediate social environment and employed only conversational methods. By contrast the sixth grade curriculum included the geography of Germany and Austria plus a drama written by the students and employed both conversational and reading methods. Thus the first graders probably had several advantages over the sixth graders in learning German pronunciation, due to higher motivation, the situational relevance of the curriculum and the emphasis on conversation.

According to the results of a pilot study done at the University of Missouri (Larew, 1960), seven-year-old subjects achieved the highest mean score on an instrument designed to determine the pupils ability to reproduce Spanish phonemes

articulated by the teacher. There were just ten subjects in each age category seven through eleven and fourteen years. They were selected on the basis of age, intelligence and absence of experience with a foreign language. After the instrument was designed, four Spanish lessons were planned and taught before the articulation test was administered (Larew, 1961, p. 204). If the ability to articulate Spanish phonemes is the major criterion for selecting an age group to begin the study of Spanish, this study would also dictate an early start.

Several studies, however, appear to question an extremely early start. They will be discussed here.

A well designed study done in Hawaii (Grinder, Otomo and Toyota, 1961, p. 197), compared second, third and fourth-grade children in the audio-lingual learning of Japanese as a second language. After eighty-six sessions the subjects' articulation of Japanese was assessed in terms of the degree to which they accurately recited ten short sentences constructed to elicit particularly nine basic characteristics important for effective articulation of Japanese. Empirical tests showed that the fourth-grade subjects were consistently and significantly higher in comprehension and speech production than those in the lower grades after an equal amount of time given to instruction with a standardized curriculum. According to these findings, FLES instruction should be introduced later rather than earlier in the elementary curriculum.

On the basis of their experiences with a five-year FLES program (one of the rare longitudinal studies) at the University of Chicago Elementary School, Dunkel and Pillet (1962, pp. 1, 4, 141, 144), assert that when audio-lingual skills are a major objective, third and fourth graders do better than those who begin later and that fourth graders do as well or better than the third graders in the oral skills. This is probably due to greater maturity, better adjustment to school routines or similar characteristics that make the fourth graders somewhat better able to profit from formal classroom work. Furthermore, they do not recommend an earlier beginning for a program such as theirs was (language mastery as the goal of the program) and conclude that much better ways are needed to determine when the student is ready for FLES.

There may be some question as to the generalizability of these findings, however, since the school in which this study was conducted had a selective admissions policy. Roughly fifty per cent of the pupils were children of University of Chicago faculty and staff members. The median IQ was always in the 130's, very few IQ's were below 110 (though it must be mentioned that IQ alone is not generally considered a very reliable indication of a student's ability to learn a foreign language, Carroll and Sapon, 1959, p. 22; Pimsleur, 1966, p. 1). The bulk of these elementary school pupils entered University High School. Over ninety per cent of its graduates attended college and large numbers continued into graduate school.

In a comparison of achievement by grade done in the well known Pennsylvania study (Smith, 1960, p. 221) students in the earliest grade achieved significantly less than students who began later. The consistent low placement of earlier grades led to an examination by grade which showed the eighth grade lowest. In seven of the eight analyses, significant differences between grades existed at the .01 level of confidence. Although this reflects the subjects performance on reading and listening tests only, the results may have implications for this study. (It might be pointed out that one of the chief objectives of the Pennsylvania study was to determine through wide scale research with Pennsylvania high school students, the relative effectiveness of three different teaching strategies: 1) traditional, 2) functional skills, and 3) functional skills plus grammar. Results showed after two years of instruction there was no significant difference among the three strategies in listening, speaking and writing. There was a significant difference in favor of the traditional method over functional skills in reading (Ibid. pp. 195-196).

Several studies indicate that we do not have sufficient evidence to support either of the above new points, i.e. neither an early nor a late start.

John Carroll, (1967, p. 137), in his study on the proficiency levels attained by language majors near graduation from college found that the time of starting foreign language study is strongly associated with the level of foreign language skill

attained (the younger the start, the better the performance). Nevertheless, he maintains that this is not a telling justification for the FLES movement. This would be true only if it could be shown that it is critical or necessary that the start of language study must be in the grade school. Many students found in his sample who started in high school or college did as well as students who started earlier. In the case of German the relationship was reversed--those who began in high school did slightly poorer at college graduation than those who started in college.

Another study with a mixed finding concerning this issue is that of Asher and Garcia (1969), who attempted to determine the optimal age for a Cuban child entering the United States to achieve native-like pronunciation. In comparing English sentences of seventy-one Cuban immigrants between the ages of seven and nineteen (most of whom had been in the United States about five years), it was determined that the younger the child, the higher the probability of pronunciation fidelity, but that some older children can also achieve an excellent pronunciation as well.

Several studies have attempted to compare the high school foreign language achievement of pupils with a FLES background to those without foreign language instruction in the elementary school. The results would seem to be of interest here, since there might be some indication as to the merit (or lack of it) of an early exposure to the sounds of foreign languages. Justman

and Nass (1956, pp. 122-123), report the findings of a study of the high school French and Spanish achievement of 100 matched pairs of pupils (on the basis of sex, age and IQ), who were and were not introduced to the study of a foreign language in elementary school. The results showed that pupils who received one year of advanced credit in French for their FLES experience generally obtained lower grades than matched pupils who began the study of language in high school. Pupils who were denied advanced credit for FLES (i.e. started over with their foreign language study) generally obtained higher grades than matched pupils who did not have previous foreign language study. Neither of the above differences were statistically significant.

In the case of Spanish the results were somewhat different. Pupils who received one term (not one year as in French) of advanced credit for FLES generally obtained higher grades than their matched pairs who began foreign language study in high school. Only the mean difference in final grades in Spanish 2 (i.e. the end of one year of study) was significant. Pupils who did not receive advanced credit for FLES generally obtained higher grades than their matched pairs with no FLES. Only the mean difference in final grades in Spanish (i.e. the end of one term) was statistically significant.

Thus in terms of high school achievement there was no apparent advantage to the pupils who had French FLES, and the better achievement of those who had Spanish FLES did not persist

over a period of years. It might be mentioned that advanced credit was received on the basis of an examination administered by the respective high school foreign language department and upon the recommendation of the FLES teacher. Unfortunately, no mention is made of the role of the speaking skill neither in the placement examination nor (more importantly) in the evaluation of achievement in the high school foreign language classes.

The performance of a FLES and non-FLES group of students on the Level M Modern Language Association Cooperative Foreign Language Tests, which measure all four skills, was compared by Brega and Newell (1965, pp. 433-438). Both groups were at the end of their third year of high school French and were enrolled in two separate classes, one with fifteen and the other with seventeen students taught by different instructors. The seventeen students in the one class had four years of French FLES in addition to their high school training. Only children who were above average academically were selected for this FLES program. The FLES group performed significantly better (at the .001 level) on all of the MLA tests than did the group which began French in high school. In addition the median raw scores of the FLES group exceeded the median scores of the national norms on all four MLA tests. This was especially evident on the speaking test.

These results should not come as a surprise, however, since in terms of total hours of training the FLES group far exceeded the non-FLES group. Furthermore, the mean IQ of the

FLES group was also significantly higher. There was significant correlation (beyond the .01 level) between IQ and MLA test scores for the non-FLES group, but no significant correlations were found for the FLES group.

Two years later these same writers (Brega and Newell, 1967, pp. 408-411) presented an extension and refinement of their earlier study referred to above. In this experiment they attempted to overcome problems which limited the generalizability of their earlier findings--significant differences in IQ, small numbers of students and different teachers for the two groups. By means of an anlysis of covariance statistic all mean raw scores for the MLA tests were adjusted to account for differences in IQ between the two groups. The findings of this study were in agreement with those presented in their earlier study--significant superiority of the FLES group in all four skills as measured by the MLA battery of tests.

Although the subjects of this study who had an earlier exposure to French (third grade) out-performed those who began later (seventh grade), it must again be pointed out, that the difference in total time of foreign language training should not be overlooked. The FLES group met eighty minutes weekly in the elementary school. Relative to the performance in speaking, which is of primary concern here, this difference would be compounded since the first two years of FLES were spent exclusively on listening and speaking.

Following is a summary of another study (Vocolo, 1967, pp. 463-469) which focused on the effect of a FLES program in French on later performance in the same language in high school. Two groups were enrolled in intermediate French. One group had had a four year FLES sequence in grades five through eight, and the second group had had the usual one year of elementary French. They were matched on the basis of intelligence, sex, grade point average and instruction received (i.e. FLES vs. one year of high school French). Results of the MLA Cooperative French Test scores at the end of their intermediate French course found the FLES group had achieved significantly better scores in all skills except reading. In speaking the difference was significant at the .001 level.

In evaluating these results one must again caution against hasty conclusions without taking into consideration a possible difference in total time of exposure to the foreign language preceding the intermediate French experience (four years of FLES as compared to one year of high school instruction) as well as the probable difference in emphasis on and time spent with the speaking skill.

Finally, an investigation will be looked at which offers some insight regarding the problem of measurement, namely the evaluation of intonation patterns. Goodman (1952, p. 70) indicates that for the teaching of intonation patterns to foreign language students and speech clinic patients, one can concentrate

on the imitation of the frequency (pitch) or duration (length) aspects of the primary or secondary stress portion of a phrase with reasonable expectations of a good performance on the whole phrase. This would seem to indicate that the measurement of intonation patterns is feasible for the present study, since German does employ pitch in its stress system (Wängler, 1966, pp. 4, 21).

As a matter of interest, the policy of the Modern Language Association will also be referred to here. According to a document entitled: "Foreign Languages in the Elementary School: A Statement of Policy" (1961, p. 1), this organization regards the years from four to eight as very favorable for beginning the learning of a second language, for it is believed that children can mimic the speech sounds and intonations accurately and can learn language patterns readily. The Association believes that, since FLES is an essential part of the long sequence needed to approach language mastery and since children imitate skillfully and with a few inhibitions in the early school years, the primary grades are the ideal place to begin language learning.

In summary then, it might be said that the related literature shows widely divergent views as to the optimum time to begin foreign language study in our schools.

CHAPTER III

PROCEDURES

Population and Sample

The subjects for this study were 210 randomly selected elementary school children, thirty from each grade, kindergarten through grade six. They were attending Tremont Elementary School in Upper Arlington, a suburban community of 40,000 located immediately northwest of Columbus, Ohio. The fifteen boys and fifteen girls from each grade were selected from the April, 1971, roster of students according to a random table of numbers (A Million Random Digits with 100,000 Normal Deviates, 1955). Thus subjects were of a similar socio-economic background and from the same geographic area. Subjects who had speech defects or who had any previous experience with or exposure to any foreign language other than popular children's programs such as "Sesame Street," "Misterrogers' Neighborhood" or "Lucy's Toyshop" were excluded from the study. Numerous subjects became inelligible due to previous FLES experience, language use in a bilingual home, etc. Out of the total of 791 pupils enrolled in Tremont School at the time the sample was selected, twenty neurologically handicapped children and one sixth grade class were not included in the study.

This sixth grade class was receiving instruction in French twice weekly for a total of 80 minutes a week. This obviously put them into the category of having previous exposure to a foreign language. The randomness of the selection was not affected, however, since pupils were placed into classrooms on a random basis. The only exception to this might have been a rare request to the contrary by a parent or teacher.

According to Mr. Walter Heischman, Superintendent of Schools, Tremont School was considered representative of the seven elementary schools within this homogeneous district. A brochure entitled "Upper Arlington Schools" prepared in 1971 by the Superintendent's Office (1950 North Mallway) indicates that there were 9010 students enrolled in grades kindergarten through twelve (4300 in kindergarten through grade six) in October of 1971. It further states that the intellectual achievement of these students is above the national average. Standardized tests record the average IQ of Upper Arlington school children to be in the 112-115 range and indicate that students are performing at one and one-half to two years above the national norm. About eighty-five per cent of Upper Arlington High School graduates attend colleges or universities. Another five to six per cent continue their formal education with vocational or technical training.

Instrumentation

The success of this experiment depended to a very high

degree upon the ability to obtain a reliable instrument. For this reason the principal investigator explored the possibility of obtaining and using an instrument already used for this purpose. This was to no avail, however, since speaking tests such as the "MLA Cooperative Foreign Language Tests" (1965), and the "Pimsleur German Proficiency Tests" (1967), are all achievement tests constructed for students who have had considerable exposure to and instruction in a foreign language, making the utterances too long and difficult for the purposes of this investigation. Another instrument which was examined for possible use was the German speech production test written by Rebecca Valette for the Pennsylvania Project. Although it appeared on face value to be too long and difficult for children age kindergarten through grade six, it was administered to two kindergarteners, one fourth grader and one sixth grader. Rarely did the utterances of these subjects even vaguely resemble those of the model. The cues appeared to be much too long and difficult for children of this age who had not had previous foreign language instruction. This instrument would have needed major revision also because of its total length. It consisted of 52 phrases. In a telephone conversation of April 5, 1971, Professor Ralph Eisenstadt (Foreign Language Department, West Chester State College, Pennsylvania), expressed doubt as to the suitability of the Valette instruments for the age group tested in this study. He was a scorer of the subject responses to the German speech production test (there was also a French

version)used for the Pennsylvania Project and had been trained to act in this capacity by Educational Testing Service, Princeton, New Jersey. Finally, the chief consultant, Professor H. H. Wängler, University of Colorado, also advised against using this instrument under these circumstances. Thus it was decided to construct an instrument expressly for the purposes of this study. In order to answer inevitable questions such as the length of words and utterances, the number of repetitions necessary in modeling and the total length of the test, a pilot test was constructed by the writer in collaboration with the consultants and administered to 39 randomly selected subjects from grades kindergarten, one, two, four and six. It was felt that it would be sufficient to select subjects for the pilot study from alternate grades thereby making it possible to select more subjects from one given grade (and still to keep the total size manageable). Since the gap between kindergarteners and second graders appeared rather large, however, first graders were also included in the pilot study. The chief consultant agreed with the above strategy.

The pilot test consisted of two parts:

a) twenty single word utterances of one or two syllables.

b) twenty phrases which, with the exception of three, varied in length from four to six syllables. (one phrase was two syllables in length and two others were three syllables long.)

Only one critical (non-English) sound or sound cluster was embedded in a non-critical environment in each word of part one and each phrase of part two. The sounds selected for testing were those which typically cause the speaker of American English difficulty. An inventory of these sounds appears under the column labeled "sound tested" on the score sheet of the pilot test found in Appendix A.

The purpose of testing the same sound in both parts one and two of the test was to evaluate the imitation of critical sounds and sound clusters within the context of both single words and short sentences, both of which are a part of our normal speech patterns. Furthermore, it was hoped that this might supply some alternate form test reliability data.

In addition to testing the production of the individual critical sounds or sound clusters in the two parts of the test, it was planned to evaluate the subjects' ability to imitate word stress in part one and sentence stress in part two. Some thought was given to including a separate part in the instrument which would contain no critical sounds or sound clusters and which would test exclusively intonation. Upon the recommendation of the consultants this idea was abandoned on the grounds that intonation could be tested simultaneously with one critical sound or sound cluster per utterance. It was also felt that testing only intonation in one section of the instrument would be inefficient,

would unduly lengthen the instrument and, of course, total testing time. With young subjects this factor is a crucial one.

With reference to one section of the instrument testing only intonation (which would have been part two) Professor William G. Moulton, of Princeton University, had the following response in his letter of April 6, 1971.

> The one general comment I have to make concerns part two of the test, which will contain no crucial sounds or sound clusters but test only intonation patterns. This is good to build up confidence on the part of the subjects. On the other hand, I doubt that you will find out much about the subjects' ability to imitate intonation patterns, since the patterns of English and German are in all basic essentials the same. (We do not always use them in the same way, though even here there are great similarities.) If the subjects do not imitate the intonation patterns properly, my guess is that this will be because the intonation patterns accompany what are, for the subjects, strings of nonsense words. This part two can do no harm; but I doubt that you will learn much from it. If the foreign language were, let us say, French, matters would be quite different.

A copy of the pilot test was drafted by the principal investigator and sent to the consultants for comment, critique and revision. The proposal was also sent so that the aims and objectives of the project would be clear. It must be emphasized that the contributions made by the extremely well qualified and cooperative consultants were invaluable and indispensable for the success of this experiment. A resume of their experiences and qualifications can be found in Appendix G.

Relative to part one of the test, Professor Moulton made the following comment in his letter of April 6, 1971.

> The words here will need to be very carefully chosen. The word zwei is again [he had referred to this word earlier in the letter] a good example. If you wish to test merely initial /ts-/, then zwei is a questionable word to use. It tests /ts-/ in the particular position before /v/. And though the subject might be quite able to imitate /ts-/ before a vowel (Zahn etc.), he might have more trouble with it before /v/. Of course, it may be precisely the cluster /tsv-/ that you want to test. Even then, Zweck might be better so that the scorers are not confused by a drawled /ai/.

The next item of concern was, of course, the selection or construction of suitable phrases for part two of the test where the advice and recommendation of the consultants was again adhered to. Following is a quotation from a letter of April 14, 1971, from Professor H. H. Wängler, noted linguist and phonetician at the University of Colorado:

> In part two I would suggest strongly to use always clearly meaningful German sentences. Even though it obviously doesn't play any role for the young listeners it might influence the production in some strange way. Most of all, it might have an influence on the judges who cannot relate to the content of the sentence naturally enough.

Similarly in her letter of April 15, 1971, Mrs. Jenni Moulton also emphasized the importance of avoiding phrases that are not normal German utterances, and did so for several reasons. They would be very difficult for a native speaker to model because they would sound forced and unnatural. Furthermore, such utterances would be difficult for the scorers to judge, since they would

confuse him and thus impair his accuracy in judging the critical sounds and sentence stress.

Thus there was agreement on all major points. The chief consultant, a native speaker of German and an authority on pronunciation and articulatory problems encountered by American speakers of German (Wängler, 1963, 1966), did the final editing of the test items. He then modeled and recorded them at 7.5 inches per second on high quality recording equipment in the Sound Laboratories at the University of Colorado, Boulder.

Since all test items were modeled by one native speaker only, the variable of different speakers (or of the same speaker modeling differently at different times) was eliminated. This made certain that all subjects heard identical stimuli. For the same reason directions were also prerecorded.

The directions for the subjects were written by the principal investigator and recorded in the recording studio of the Listening Center in the new Dieter Cunz Hall of Languages on The Ohio State University Campus. The reader was Edith Walters Cole, a Ph.D. candidate in speech at Indiana University, Bloomington. Her experience, voice quality and diction eminently qualified her for this duty.

The total script of the directions for the pilot test can be found in Appendix B. Every effort was put forth to make the directions as clear and lucid as possible. Provisions were made for the subjects to practice several items (including English as well

as German) preceding the actual test items, in both parts one and two of the instrument.

The final editing and splicing of the two recordings (instructions and actual test items) was also done at the Listening Center at The Ohio State University. It can be said without any reservations whatsoever, that the quality of the master tape was truly excellent. This was verified by the experienced and professional staff at the listening center including the director, Mr. Gary Mann and two of his able assistants, Mr. Bill Logsdon and Mr. Jim Keckley.

The randomly selected subjects were invited to participate by letter a copy of which can be found in Appendix C. The following strategies were employed to encourage a high degree of participation from those selected.

1. The off-set process was used for duplicating the letter of invitation since it produces copies very much resembling the original.

2. Envelopes with an Ohio State University return address were used for the mailings.

3. The envelopes were addressed directly to the subjects rather than to the parents. It was felt that this would make it more personal.

4. A list of all subjects who had been selected was enclosed with each letter so that friends and classmates who were interested in doing so, could come together to the same testing session.

5. Mention was made of the fact that this project was being undertaken with the approval of the central administration of the school district.

6. It was pointed out that the study was supported by a grant from the U. S. Office of Education.

7. Round trip bus transportation to and from the testing site was provided free of charge.

8. An adult chaperone accompanied the groups on each trip.

9. Younger subjects could be accompanied by a parent, guardian, an older family member or friend, etc., if they so desired.

10. Several testing sessions were held at different times of the day so subjects could come at a time of their choice.

11. A stamped self-addressed envelope was enclosed with the letter in which the subjects and/or parents returned a simple form (a copy of which can be found in Appendix D) indicating at what time the subjects wished to participate.

12. Subjects were promised that their test scores and brief results of the study would be available at their request.

13. Since a limited number of people was involved, each invitation was followed with a telephone call by the principal investigator to confirm the time of testing and to answer any questions.

The response to the invitation was very positive in terms of both the percentage of subjects who responded as well as in terms of the enthusiasm displayed by parents and subjects alike. Of the fifty subjects who were invited to participate (ten from each age group), thirty-nine were tested making a 78 per cent response. Several subjects indicated that they would like to participate but could not due to illness, prior commitments, etc.

A bus was leased from The Ohio State University Transportation Department in order to transport the subjects to and from the parking lot of Tremont Elementary School (which all subjects attended) and the Dieter Cunz Hall of Languages on the Ohio State University Campus.

The three different testing sessions were conducted in one of the new and ultra-modern language laboratories which are part of the Listening Center complex. The laboratory had thirty-five booths all of which were equipped with remote-control Raytheon tape decks located in cabinets at the front of the room. Subjects were allowed to sit in booths of their choice, with the exception that they were asked to sit towards the front of the room, since the lab was only partially filled for each testing session. All booths were ready for use, however. The volume control had been preset so testing conditions would be equal for all subjects in every respect. There are no windows in the lab so there were no distractions of any kind. In sum, it might be said that testing conditions were ideal.

By way of introduction and preliminary routine, the following uniform procedures were adhered to at each testing session.

1. Subjects were told to relax and feel at home, because they would not need to worry about how to operate any equipment in the lab. Indeed they were even assured that they need not even touch any of the buttons on the panel in their booths.

2. It was explained that subjects would not be required to spend long periods of time at one task, and that there would be a pause for rest and relaxation between activities.

3. It was pointed out that if anyone encountered any difficulties, he should simply raise his hand, and someone would assist him. (In addition to the principal investigator who was serving as a proctor and supervisor, one of the full-time lab technicians was on duty to operate the controls at the console and to remedy any technical problems.)

4. Subjects were urged to pay very close attention to the directions they would hear on the recording.

5. The proper way to put on the head-set was demonstrated. As a double-check, however, the proctor went around to each subject, adjusted the head-set for size and comfort, and then placed the microphones (which were mounted on the head-set) close to and directly in front of each subject's mouth.

6. Subjects were asked on the "all-call" to raise their hands if they could hear the proctor, to ensure that the sound out-put was functioning properly.

7. As a final step the numbers which had been randomly assigned to each subject were rehearsed. This served two functions. It gave the subjects a chance to hear themselves speak over the amplified headsets, and secondly, they were more prepared to give their number when asked to do so very early in the exercise. In an effort to guarantee anonymity, subjects were asked to give a number rather than their names. Large slips of paper with the numbers on them were placed in each booth. Numbers were randomly assigned with the exception that no number larger than twenty was assigned to kindergarteners. Furthermore, these numbers were practiced with kindergarteners before they put on their headsets to make sure they knew them. There appeared to be no problems in this respect. The purpose of assigning the numbers randomly was so that the judges could not identify the age of the subject by the number which they heard when they scored the subject responses.

8. Steps four, five and six were repeated between parts one and two of the test.

Although the laboratory equipment was just a few months old, the adage "whatever can go wrong will" still held true. One subject could not hear the cues shortly after the start of one session so she was placed in a different booth. Fortunately,

this was the only technical difficulty encountered.

During the actual testing session, each cue was heard two times by the subjects. The signal for the response was the bright illumination of the red button on the control panel in the booth. This had, of course, been pointed out in the instructions. This button remained illuminated during the entire eight seconds the subjects had for their response. The timing of this light was synchronized with the remotely controlled tape decks so that the respective tape recorders were running only when the lights of the corresponding booths were on. The signals to achieve this had been put on the second (half) track of the master tape. The German stimuli for the subjects were, of course, on the other half track. The net result was that only subject responses were recorded, omitting the stimuli of the native speaker. It is estimated that this process of omission saved the judges approximately half the time it would take them to score the responses otherwise. Since all of the judges were sent a copy of the master tape and instructed to become thoroughly familiar with it, it is not estimated that this process hindered their scoring accuracy. Furthermore, it will be noted that the score sheet contained a written version of the stimuli, with the sound to be scored underlined. The phonetic transcription of this sound was also printed on the score sheet (Appendix A).

The importance of the scorers' familiarity with the master tape (subject stimuli of the native speaker) was underscored by

Professor Wängler in his letter of April 28, 1971, when he said: " . . . all the judges need it, and special instructions to listen to it too."

After each testing session a new magnetic tape was placed on the tape decks that had been used. After the final testing session was completed, the entire set of recordings was put through an equalizer to boost the record level and to filter out some extraneous noise. The individual subject response tapes were taken in random order (to reduce judge bias relative to the age of the subjects) and dubbed on five inch reels of magnetic tape. These were sent to the judges accompanied by the master tape and score sheets.

There was a considerable amount of debate as to the scale to be employed in evaluating subject responses. Opinions focused on a four to five-point scale. On the five-point scale scores would be assigned according to the following criteria:

0 - no response
1 - not acceptable
2 - acceptable
3 - near native
4 - native.

In a letter of April 2, 1971, Ralph Eisenstadt (mentioned earlier in relation to the Pennsylvania Project) wrote a comment in response to the four point scale discussed in the proposal: "The scale you indicate seems quite in order and I see no reason to change it before the pilot." The four point scale differs from the five-point scale in that "near native" and "native" are

lumped together. In his letter of April 6, 1971, Professor Moulton explained that on a theoretical basis his first reaction was that one should have a five-point scale. Mrs. Moulton, who has a great deal of experience as a scorer for Educational Testing Service favored the four-point scale. She thought it allowed not only faster and more accurate scoring, but also far better judge agreement. Her experience with ETS scorers was that on the five-point scale they found 1, 2, and 5 relatively easy to assign, but wasted a lot of time worrying about the distinction between 3 and 4, and showed little convergence. Professor Wängler also expressed his support for a scale wider than two points (used in the Pennsylvania Project) in his letter of April 1, 1971, when he said: "'Acceptable' against 'unacceptable' brings a pseudo-exactness of sorts into the results. Your results will <u>look</u> more impressive but will by no means <u>be</u> more exact. I am for at least a four-step rating scale; . . . " Professor Tom Wilkey of the Math-Statistics Department at The Ohio State University pointed out that in terms of data analysis, a four-point scale would be suitable, because if it did not turn out to be practical and usable, one could easily collapse to a two-point scale where 0 and 1 would be "unacceptable" and 2 and 3 would be "acceptable."

Thus the sentiment was clearly in support of trying <u>at least</u> a four-point scale for the pilot test. Since the four-point scale was suggested as a minimum it was reasoned that it might be a sound decision to experiment with a somewhat wider scale

for the pilot test on the grounds that one could easily collapse to a four-point scale for the final test if the five-point scale proved to be unusable. Furthermore, it seemed much more feasible to expand from a five-point to a six-point scale (if the judges felt this was necessary on the basis of their experiences scoring the pilot test) rather than to expand from a four-point to a six-point scale in the event the four-point scale proved to be very inadequate. In short, the five-point scale was considered a better middle ground from which to digress. Therefore the five-point scale was used and the score sheet (which appears in Appendix A) included the sixth step in the event the judges wanted to experiment with it. Extra score sheets were provided for this purpose.

The decision on how to rate word and sentence stress was not so difficult, since it appeared to be merely a question of whether to employ a simple two-point scale (right-wrong) or a three-point scale. In response to the suggestion of using the former, Dr. Paul Pimsleur at the State University of New York, Albany, (formerly Director of The Ohio State University Listening Center) responded in a letter of April 16: "You might consider stretching the scale for judging intonation from two to three points (good, fair, poor)." On an experimental basis the pilot test was scored on the 5-point scale. The score sheet was designed to entertain an even wider scale in the event the judges wanted to experiment with this possibility.

In a telephone conversation of May 16, 1971, Wängler confirmed that a three-point scale would be satisfactory and requested that the format of the score sheet for the final test (see Appendix E) should provide for the three-point scale in evaluating word and sentence stress.

The response to the five-point rating scale for evaluating the individual sounds and sound clusters was also quite positive. After having scored the subject responses from the pilot test, the judges reacted as follows. Mrs. Moulton in her letter of May 4, 1971, said: "I am in favor of using the five-point scale, and I have scored the tapes on this scale, using the directives you suggested:

0 - no response;
1 - not acceptable;
2 - acceptable;
3 - near native;
4 - native."

In his letter of May 10, 1971, Professor Wängler (speaking also in behalf of Mrs. Wängler) remarked: "Basically the five-point scale worked fine with us. Sometimes we missed the possibility to give a grade between one and two. Between 'not acceptable' and 'acceptable' we could have used something like 'barely acceptable.' I would definitely not cut down on the number of choices."

Since Mrs. Moulton did not participate in the final scoring, there seemed to be no reason not to accommodate the other two scorers. So the six-point scale was used for the final test.

With reference to scoring the subject responses, the question was raised as to whether judges should do the scoring simultaneously at one given location. Mrs. Moulton addressed herself to the question as follows in her letter of April 15: "I don't think it is at all necessary for the scorers to work simultaneously. In my experience scorers work better, more accurately, and hence more economically when working by themselves. . . . (ETS now farms the scoring out to individual scorers.)" Wängler reinforced this idea in his letter of April 28, 1971: "Mrs. Moulton is right, of course; I would go even one step farther, judges shouldn't work together." Thus the three native speaking judges were each sent a separate recording of the master tape and the subject response tapes: Professor H. H. Wängler, Boulder, Colorado, Mrs. Ilse Wängler and Mrs. Jenni Karding Moulton, Princeton, New Jersey.

After the pilot test was scored, it was revised and the final version of the instrument was drafted on the basis of the pilot test results. These will be discussed in detail in Chapter 4. Revisions were made by the principal investigator and sent to the consultants for approval and/or alteration. Where there was disagreement (which was mild) the advice of the chief consultant, Professor Wängler was adhered to.

In summary it might be said that the pilot test was an extremely worthwhile experience, not only in terms of the information it yielded relative to the specific test items, but also

relative to innumerable logistical details. As a matter of fact, this writer would even go so far as to say that for an experiment of this nature, a pilot test is not only highly desirable, but an absolute prerequisite.

The same basic procedures were followed for the final test as for the pilot test. Naturally changes and improvements were made on the basis of the experience and knowledge gained with the pilot test. These will be discussed here.

1. Sample size: The final test was administered to 237 subjects. As the total indicates, more subjects were tested than were actually needed (210 or 30 from each of seven age groups). It was felt that over-subscribing was necessary in the event of a technical failure of the equipment or the accidental erasure of a recording for example, which would have necessitated later testing sessions. Only one recording was not usable. Apparently this subject moved his microphone too close to his mouth after it was adjusted for him, causing distorted responses. Needless to say, the responses of the first thirty subjects selected in each age category according to the random sampling procedure were subjected to the data analysis.

2. The letter of invitation was printed on official Ohio State University letterhead paper.

3. Free round-trip transportation was again provided but with the larger numbers of subjects involved, the assistance of two chaperones was acquired. One accompanied and supervised

the children on the bus and made sure that no one was unintentionally left behind. A list of subjects had been prepared in advance according to the time subjects had indicated they would participate. Thus an accurate account was kept of which subjects were tested at which session. For the sake of efficiency the testing schedule was arranged so that while one group was being tested the bus was returning to Tremont School for the next group. The second chaperone acted as what might be called a "receptionist" at the school parking lot making sure no one arriving for the testing session would leave (in the case of an early arrival) simply because the bus had not returned from Dieter Cunz Hall of Languages. Likewise this chaperone, who was acquainted with many of the students as well as the research project stayed with those children who did not have an immediate ride home when the bus departed with the subjects for the next testing session. This arrangement proved to work out very smoothly.

4. Again due to the much larger number of subjects, two language laboratories were used instead of just one. The equipment in the two was identical. Since this required two proctors, the services of a second proctor were acquired. Mr. William E. DeLorenzo served in this capacity. At the time, he was a Ph.D. candidate in Foreign Language Education at The Ohio State University. His wide experiences as a foreign language teacher and with language laboratories were a great asset. He was given a

detailed list of duties he was to perform and these were discussed thoroughly in advance. In addition, he observed the other proctor (the principal investigator) during the first testing session which was held in one lab. During subsequent sessions whenever both laboratories were used, the group of subjects was divided according to age. Depending on the number of subjects present in each age category, kindergarten through grade three went to one lab, for example, and the older subjects to the other. The two rooms were directly adjacent to each other. To avoid any possibility of one proctor having a significant influence on a given age group's performance, the two proctors alternated age groups from one session to the next.

5. Subjects were allowed to sit in alternate booths only. This was done to decrease the amount of test leakage and also to avoid possible distractions from subjects in adjacent booths. Although this meant that the laboratory could be filled to only about one-half of its capacity, it was felt by all personnel involved (proctors as well as lab technicians) that this was more than worth the sacrifice in loss of space.

6. The warm-up exercises prior to the testing session were lengthened to give the subjects more time to get adjusted to their surroundings and to hearing their own voices on the amplified head sets. They were asked to repeat their names several times and to count in unison from one to five.

7. Subjects were asked to identify themselves by name rather than by number, and directions were changed accordingly (see Appendix F). Several tapes in the pilot test were not identifiable because some subjects failed to give their number when asked to. Other recordings had duplicate numbers. The importance of remaining anonymous at this point did not appear to be as crucial as being able to identify the recordings. To further help alleviate this problem a written record was kept to see which booth was occupied by which subject for each session. Thus if a name was missing from a recording (which did happen in several cases) it could be identified by checking to see who sat in that given booth in the sessions immediately prior to and following that session.

8. In order to increase motivation and interest, subjects were told that this was a little contest among grades kindergarten through grade six, to see who could get the best score.

9. Rather than removing the reels from the tape deck after each testing session, the subject age group in each row was randomly changed. This reduced the problem of possible judge bias, relative to age and was also more efficient, since changing tapes for each group tested would have necessitated waiting periods. Some blank tape was left between recordings. Pimsleur in his letter of April 16, 1971 had warned: "I'm sure you'll randomize the order in which tapes are scored, and will disguise the identity of the students. Still, I'm (only mildly)

concerned that the scorer may know by the voice quality which group he belongs to and score the young ones more leniently . . . . "

The principal investigator can vouch for the difficulty of identifying the age of subjects on the basis of the recorded utterances, even when trying to do so! In the interests of saving time and avoiding frustration for the scorers, the names of the subjects were typed on the score sheets (see Appendix E) before they were sent. As the names were typed, they were checked off a list which had been prepared according to the grades of the subjects. It was for all practical purposes impossible to identify the grade level of the subject on the basis of voice quality. In many cases the miscalculation represented four grades, for example. In this respect it would have been advantageous to use numbers for identification because: 1) the numbers would have been shorter than names, 2) scorers would not have been able to identify the sex of the subjects by their name and consequently perhaps not be in as good a position to deduce the age of the subjects.

10. Stimuli were heard three times rather than two times as in the pilot test. The reasons for this will be discussed in greater detail in the next chapter.

Data Analysis

Three variables were analyzed:

1) Judge agreement: judge 1 and judge 2.

2) Grade level: kindergarten, 1, 2, 3, 4, 5, 6.

3) Sex: boys and girls.

Judge agreement was checked on each item, on the four individual parts of the test and on the total test, by subjecting the test scores assigned by each judge to the Mann-Whitney-U (MWU) and the Fisher Student t (FST) tests. (The MWU was used since the scores assigned by the judges are not clearly interval scale data. They represent a value on a scale rather than interval data. The t Test was run to provide a basis of comparison. There was very high agreement between the two.) In addition, two correlation coefficients were calculated on the basis of total scores assigned to each subject by each judge: Pearson Product-Moment Correlation Coefficient and Spearman Rank-Order Correlation Coefficient.

Differences among grade levels and the two sexes were analyzed by using the same two tests (MWU and FST). But the data used for these analyses were an average of the scores assigned each subject by the two judges for each item, for each of the four parts of the test and the total test. These two variables (grade level and sex) were analyzed on the basis of scores achieved by all subjects (Appendix I) and then on the basis of scores achieved by the top ten subjects from each grade (Appendix J).

Test reliability was calculated by a Pearson Product-Moment Correlation Coefficient on scores achieved on odd and even numbered items. This split-half correlation coefficient was then adjusted using the Spearman-Brown Prophecy Formula.

To ascertain which items were best discriminating among the various grade levels, an item-by-item analysis was done with the MWU and the FST tests comparing each grade with every other grade.

Mean scores and standard deviations were calculated by grade level on each of the four parts of the test as well as total scores achieved by all subjects, and by the top third of the subjects from each grade level. Mean scores (parts 1A, 2A, and total test) were plotted accordingly.

After calculating the means on each item in part 1A and 2A of the test, the items were placed in rank order according to difficulty.

Personnel

Much of the success of this study must be attributed to the outstanding personnel who performed various duties in relation to this project. Their unique qualifications (which can be found in Appendix G) speak for themselves. Following is a list of these people and a summary statement of the capacities in which they served:

1) Dr. Edward D. Allen: Project Director.

2) Dr. L. O. Andrews, Consultant.

3) Roy Carlson: Technician.

4) Edith Walters Cole: Reader for directions of the tests.

5) William E. DeLorenzo: Language laboratory proctor (final test).

6) Rudolfo Garcia: Chaperone (final test).

7) Dr. Gilbert Jarvis: Consultant.

8) Jim Keckley: Technical Supervisor of the Listening Center.

9) Bill Logsdon: Recording technician.

10) Gary Mann: Acting Director of the Listening Center.

11) Jenni Karding Moulton: Consultant for both the pilot and final tests; scorer for the pilot test only. (Mrs. Moulton had agreed to be a scorer for the final test also, provided that the subject response tapes would reach her no later than May 27. This was dictated by previous commitments. At the time she was originally contacted, this appeared to be no problem. Unfortunately this time schedule could later not be met.)

12) Dr. Paul Pimsleur: Consultant.

13) Dr. H. H. Wängler: Chief consultant, scorer and model of German stimuli.

14) Mrs. Ilse Wängler: Scorer.

15) Ardis Wipf: Chaperone.

CHAPTER IV

RESULTS

Pilot Test

The results of the pilot test indicated the following:

1. Part one of the test could possibly have been used in its original format for the final test. (It consisted of twenty words which were one and two syllables in length. Each word had one critical sound or sound cluster embedded in a non-critical environment.)

2. Part two of the test had to undergo major revisions before it could be used again because the imitations were too distorted to merit scoring them. (This part of the test consisted of twenty utterances, with few exceptions, four to six syllables in length. Each utterance, as in part one, had one critical sound or sound cluster embedded in a non-critical environment.) The probable reasons for this were: a) Many of the utterances were too long. Four syllables appeared to be a reasonable maximum length of utterances. This led to revisions such as: "Sie haben heute Pech" (6 syllables), to "Sie hat Pech" (3 syllables); "Zwischen ihm und uns" (5 syllables), to "Zwischen ihnen" (4 syllables.) b) The sequence of sounds, including non-critical sounds, appeared to have a great deal of influence on how well

subjects could imitate critical sounds. Examples of this were: "Die Sonne Scheint," with two slightly different fricatives in two successive words and "Uwe muss üben," where the labiodental, voiced fricative /v/ in "Uwe" plus the bilabial, voiced stop /b/ in "üben" caused considerable difficulty (note that both sounds have labial elements). The sequence of the two different vowels immediately preceding these two sounds may have complicated the utterance as well. It appeared that this sequence of sounds unduly complicated the utterance and was not a valid test of the subjects' ability to produce the long German ü-sound. Of course it is of interest to see how well students can imitate sounds in various phonetic contexts, but to test all critical sounds in numerous contexts would require an impractically long instrument.

3. Two modelings of the utterances in part two of the pilot test were probably not sufficient.

Since the above three findings seem to be closely related (they all deal with test content and format), the writer will address himself to them here. Under the circumstances there seemed to be several alternatives at this point.

a) Drop part two of the instrument (phrases) and lengthen part 1 (single words).

b) Shorten all phrases in part two to a maximum of three to four syllables and keep the basic format of the test the same.

c) Simplify the sound pattern of the utterances.

d) Model the utterance three times.

e) Place all sounds to be scored in a primary stress position, as most of them already were. Although there was slight disagreement on which steps to follow (mostly onsteps a and b) it was considered in the best interests of this study by the principal investigator and the chief consultant to implement the combination of steps b - e. Concerning the choice between a and b, it was felt that there should definitely be some attempt to include utterances of more than one or two syllables in length. Using phrases appeared to be the best solution, because as single utterances in natural German speech, short phrases would no doubt be more common than long words. Secondly, as words get longer it gets more and more difficult to find common ones that meet the prerequisite of only one critical sound or sound cluster per utterance. There seemed to be much more flexibility with phrases. Furthermore, using phrases rather than single words also made part two distinctively different in nature from part one, which was revised to include one-syllable words only. With part one consisting of one-syllable words it could be said with assurance that the length

of the utterance was not the main factor in the performance of the subjects, but rather that it was the relative difficulty of the utterance. At any rate, the only way to make them any shorter would have been to isolate the individual critical sound or sound cluster completely. Both the pilot test and the final test showed that it is not necessary to make the utterances that short.

As for the number of modelings, a brief post-test questionnaire revealed that about one-third of the subjects would have preferred to hear the stimuli three times. It was felt that the third modeling was in order on this basis. Practical classroom experience also dictates that modeling a totally new utterance for students three times is not excessive. Although two modelings might have been sufficient for a few of the items testing vowels in part 1, it is doubtful that one could justify changing directions on these grounds in any part of the test and possibly confuse the subjects by telling them they will be hearing two modelings on one part and three in another.

4. Subjects tend to make the same types of errors in a target language as children do in their native language. Mase (1962, p. 605), lists three of the following four types of errors as among the most common ones. (Examples of errors subjects made on the test are given in each case.)

a) Substitutions: American l- and r- sounds for the German variety.

b) Omissions: Omitting part of an affricate- /faɪfə/ for /pfaɪfə/.

c) Additions: placing a vowel between two consonants as: /kənɪf/ for /knɪf/.

d) Inversions: /gəsɪçst/ for /gəsɪçts/.

5. The influence of the native language on the target language was a definite source of interference for the subjects. The examples given above under "substitutions" illustrate this point.

6. It is possible for a pair of judges to evaluate the production of critical sounds and sound clusters on a five-point scale and to do so with a fair degree of inter-judge agreement. Several tests were run to check the extent to which the two judges agreed on the evaluation of the subject responses. One of these was an item by item correlation coefficient on part 1 of the test (see Appendix A), as listed in Table 1.

Eighty-five per cent of the items show a correlation of above .62 and fifty-five per cent correlate at .71 or higher. Item 11 has a correlation of .00 because judge one did not vary his score of one on any of the subjects. (Note that this does not include the scoring of intonation.) A rather low correlation resulted in item 10, "Baum" with the dipthong as the critical sound. The correlation on item 13 (l-sound) is also somewhat low.

Table 1: Pilot Test, Correlation of Judge Agreement on Part 1A

Item Number	Correlation Coefficient	Item Number	Correlation Coefficient
1)	.72	11)	.00
2)	.86	12)	.77
3)	.88	13)	.56
4)	.75	14)	.81
5)	.78	15)	.73
6)	.69	16)	.68
7)	.65	17)	.64
8)	.72	18)	.71
9)	.62	19)	.80
10)	.37	20)	.65

Judge variability was also checked by running the Mann-Whitney U Test, which does a statistical analysis of non-parametric data. It confirmed that the differences between the two judges on item 10 were significant at the .05 level. The difference on item 11 was significant at the .005 level. This shows that there was a significant difference in judge agreement on just two of the twenty items relative to the evaluation of sounds and sound clusters.

7. Judge agreement was not as high (judged on the basis of correlation coefficients) on word stress. An item by item correlation coefficient showed the results tabulated in Table 2.

There is a correlation of .60 and higher on just sixty per cent of the items, and a correlation of .71 and above on just thirty-five per cent of the items. No doubt the wide scale (5-point) on which responses were rated was partially responsible for this. The Mann-Whitney U test for judge variability showed that judge variance was significant on only one item relative to word stress.

8. On an item by item analysis, a number of statistically significant differences were found relative to scores achieved by each grade level. In relation to word stress, just the last item produced any significant difference. Since the sample from each grade level was quite small (kindergarten (K) -4, first grade -7, second grade -5, fourth grade - 8, sixth grade - 7),

Table 2: Pilot Test: Correlation of Judge Agreement on Part 1B

Item Number	Correlation Coefficient	Item Number	Correlation Coefficient
1)	.60	11)	.50
2)	.53	12)	.50
3)	.81	13)	.59
4)	.73	14)	.68
5)	.46	15)	.82
6)	.63	16)	.81
7)	.71	17)	.63
8)	.33	18)	.56
9)	.73	19)	.66
10)	.49	20)	.87

it was not possible to draw any conclusions on the basis of these results.

Final Test

Attention will now be turned to the results of the final test in which 210 subjects were tested, fifteen boys and fifteen girls from each grade level kindergarten through grade six. Following is a list of variables that were analyzed.

1) Judge agreement: judge 1 and judge 2.

2) Grade level: kindergarten 1, 2, 3, 4, 5, 6.

3) Sex: boys and girls.

A detailed explanation of the statistical analyses utilized can be found in Chapter III, in the section discussing data analysis. In reporting the results only those differences which were statistically significant at the .05 level or higher will be given. Others will be labelled NSD (no significant difference).

Variable 1: Judges: The purpose of this analysis was to see how well the two judges agreed on the respective scores they assigned to the subjects. The following table shows the item by item results of the Mann-Whitney-U(MWU) and the Fisher Student t (FST) tests relative to judge variability on part 1A (critical sounds and sound clusters embedded in single words) and on part 2A (critical sounds and sound clusters embedded in phrases.)

Table 3: Final Test: Judge Variability on Parts 1A and 1B of Final Test

Item Number	Sound Tested	Level of Significance Part 1 MWU	Part 1 FST	Part 2 MWU	Part 2 FST
1)	[i:]	NSD	NSD	NSD	NSD
2)	[e:]	NSD	NSD	NSD	NSD
3)	[o:]	NSD	NSD	NSD	NSD
4)	[ɔ]	.04	NSD	NSD	NSD
5)	[a]	.0001	.001	NSD	NSD
6)	[ø:]	.004	.005	NSD	NSD
7)	[œ]	NSD	NDS	.007	NSD
8)	[y:]	NSD	NSD	NSD	NSD
9)	[Y]	NSD	NSD	NSD	NSD
10)	[a^{U}]	NSD	NSD	NSD	NSD
11)	[ç]	.02	.03	NSD	NSD
12)	[X]	NSD	NSD	NSD	NSD
13)	[l]	NSD	NSD	NSD	NSD
14)	[R]	NSD	NSD	NSD	NSD
15)	[ʁ]	NSD	NSD	NSD	NSD
16)	[pf]	NSD	NSD	NSD	NSD
17)	[kn]	NSD	NSD	NSD	NSD
18)	[ʃt]	NSD	NSD	NSD	NSD
19)	[tsv]	NSD	NSD	NSD	NSD
20)	[çts]	NSD	NSD	NSD	NSD

This table reveals that in part 1A both tests showed judge variability was statistically significant on items 5, 6 and 11. In addition, the MWU showed a significant difference on item 4. It was thought that perhaps an unusually wide range of scores on these four items might account for judge variance. This was not the case, however, as scores ranged from 0 to 4 on all of these items and seven others had a narrower range—from 0 to 3.

These results also show that on this part of the instrument, judge agreement was more difficult to achieve on vowel sounds than on consonants and consonant clusters. Perhaps this is due to the fact that both quality and quantity are factors affecting the relative score assigned by the judges on vowel sounds. This is not the case with consonants where the judges used basically one criterion in assigning scores. In his letter of July 20, 1971, Wängler wrote in behalf of the two judges:

> In vowels, quality and quantity realization played the most important roles. There is no possibility to objectively determine the percentage of involvement of either one in our decisions. You simply judge the sound as a whole in the given surroundings. My personal impression is, though, that we punished quality deviations more severely than quantity deviations . . . For consonants, my best educated guess is that "precision" plays the major role. Of course, in different words the same consonant needs a different degree of precision for the native listener, even if it occurs in the same position, the syllable with the main accent, for example . . . In consonant clusters the precision of the first part seemed to be more important for a better "grade"

than the following one. In /tsv/, for example, a fairly high degree of labialization could be tolerated (/tsw/), if only the /ts/ was clear. On the other hand, a clear /v/ didn't do much good if /ts/ was not acceptable. Again, the word in question had a great influence. In "zwischen" every German tolerates more labialization than in "Zweck," for example.

In part 2A, item 7 was the only one on which judge variability was statistically significant, and only as measured by the MWU test. It is of interest to note that this was not one of the sounds on which judge variability was significant in part 1A of the test. Furthermore, it came as somewhat of a surprise that there were four items (according to the MWU test) which showed judge variability at a statistically significant level in part 1A and just one such item in part 2B. One might think that judge agreement would be more difficult to achieve when evaluating a critical sound or sound cluster embedded in a phrase rather than in a single word. There would seem to be more distractions from the sound to be scored in a longer utterance. This is apparently not the case. Perhaps the judges were able to score part 2A of the test with greater agreement because they were now hearing a given voice saying similar sounds for the second item. It is also possible that the phrases sounded like more natural speech utterances than single words and that this had an effect on scoring. All sounds scored appeared in a primary stress position in both parts of the test, so a difference in stress was probably not responsible for this phenomenon. Since all analyses except, of

course, judge variability were made on the basis of average scores assigned the subjects by the two judges, their disagreement on just a few items should not have affected the results of the test.

A word of caution is in order here relative to any comparisons made between part 1A and 2A of the test. In his letter of July 20, 1971, Wängler pointed out that one of the dangers of making such comparisons is the fact that the sounds appear in different surroundings in the two parts of the test. The phonetic context may be similar but not identical. The amount of stress may also vary. In addition the length of the utterances is different. This is not to say that comparisons cannot be made, but rather that when they are, various factors must be kept in consideration.

In a composite analysis of judge variability on all items in part 1A of the instrument separately and all items in part 2A of the instrument separately, neither the MWU nor the FST showed any statistically significant differences.

An analysis of judge variability on part 1B (word stress) and part 2B (sentence stress) with both the MWU and the FST tests showed the following:

1) NSD on any individual item in either part.

2) NSD on either part analyzed as a whole.

On an analysis of judge variability on all four parts of the instrument combined, neither the MWU nor the FST test

showed any statistically significant differences between the two judges.

Finally, the Pearson Product-Moment Correlation Coefficient calculated on the basis of total scores assigned to each subject by each judge was .96 and the Spearman Rank Correlation Coefficient was .95.

In summarizing the discussion on this variable it might be said that only five individual items on both parts 1 and 2 showed any judge differences to be statistically significant. Considering the four individual parts of the test separately and the instrument in its entirety, there was no significant difference among the scores assigned by the two judges. Furthermore, the correlation coefficients were very high. In essence these two judges passed a rather rigorous inquiry into their agreement on the evaluation of subject responses.

Variable 2: Grade Level: (Part 1A: Sounds and Sound Clusters in Single Words.) The purpose of this part of the analysis was to determine if there was any statistically significant difference in the scores achieved among the seven grade levels who participated in the study: kindergarten through grade six. An item-by-item analysis comparing each grade level with every other grade level produced the following results. All differences significant at the .05 level or higher are reported.

In the following tables the grade or grades appearing under the column labelled "higher groups" achieved a significantly

Table 4: Part 1A: Sounds Embedded in Single Words--Significant Differences Among Grade Levels on Each Item

Item Number	Sound Tested	Higher Groups	Lower Groups
1	/i:/	NSD	NSD
2	/e:/	4,6	K
		1,3,4,6	2
3	/o:/	NSD	NSD
4	/ɔ/	4,5,6	K,1,2
5	/a/	4	2
6	/ø:/	3	1
7	/œ/	4	2
8	/y:/	4	K,1
		6	K,1,2,5
9	/Y/	5,6	K,1
10	/aU/	3,5,6	1
11	/ç/	1	4,6
12	/X/	1,3,4,5,6	K
		4,6	2
		6	3
13	/l/	NSD	NSD
14	/R/	NSD	NSD
15	/R/	4,6	1
16	/pf/	5	2
		6	K,1,2,3
17	/kn/	6	K,2
18	/ʃt/	2,3,4,5,6	K
		2,3,5,6	1
		6	2,3,4,5
19	/tsv/	1,3,5,6	K
20	/çts/	5,6	K
		6	1,2,4

better score than the grade or grades in the column labelled "lower groups."

A cross tabulation by grade level and group, lower and higher, (Table 5), makes the above inventory more meaningful. It shows not only the total number of times a given grade level achieved a significantly higher or lower score than other grades, but also indicates how many times each individual grade level exceeded every other grade level. Reading across in the third row (second grade) one can see that second graders did significantly better than kindergarteners and first graders one time each. Reading down in the fourth grade column, for example, shows that the fourth grade was significantly lower than the first grade on one item and lower than the sixth grade on two items.

Table 5: Part 1A: Cross-Tabulation of Significant Differences Among Grade Levels on Individual Items

High Grade	Low Grade K	1	2	3	4	5	6	Totals
K	X	0	0	0	0	0	0	0
1	2	X	1	0	1	0	1	5
2	1	1	X	0	0	0	0	2
3	3	3	1	X	0	0	0	7
4	5	3	5	0	X	0	0	13
5	6	4	2	0	0	X	0	12
6	10	8	8	3	2	2	X	33
Totals	27	19	17	3	3	2	1	72

In analyzing this cross-tabulation it immediately becomes evident that the upper grade levels (3-6) did overwhelmingly

better than the lower grade levels (K-2). Comparing these two broad categories it is noteworthy that the second graders and kindergarteners did not exceed the 3rd through 6th graders in a single instance. Within the lower group (K, 1st and 2nd graders) the first graders were notably the best, and the kindergarteners the poorest.

Looking at the other grade levels (3-6) it becomes apparent that, with the exception of grade four, the higher the grade level the better the performance. The fourth grade earned a rank of "higher grade" on one more item than the fifth grade but the two appeared in the "lower grade" column an equal number of times. Pitting just these two grades against each other in the "high and low" columns and rows, shows that neither was significantly better than the other on any item. What is especially noteworthy among the upper grades is the stellar performance of the sixth grade. Their achievement ranked them the "high group" 33 times, more than twice as often as the nearest competitor. Furthermore, they are the low group just one time--exceeded here only by the first grade.

Certain trends have been discovered on the basis of the results on individual items. The results of a comparison of composite scores on part 1A by grade level will now be displayed and discussed.

Placing the grade levels in rank order from highest to lowest, according to Table 6, confirms the trends which emerged earlier.

Table 6: Part 1A: Significant Differences Among Grade Levels on Composite Scores

Higher Groups	Lower Groups
3,4,5,6	K
4,5,6	1
3,4,5,6	2
6	3

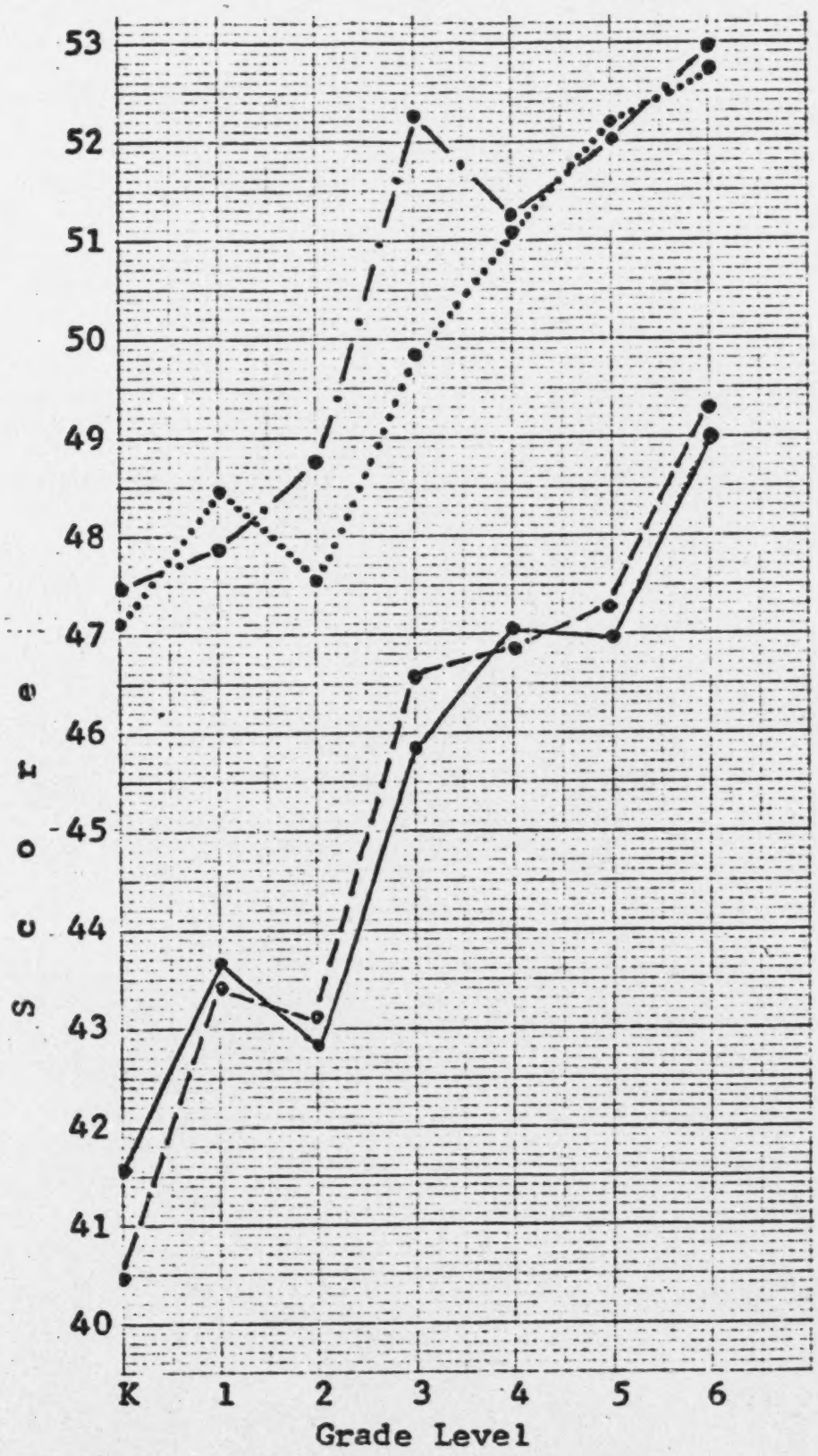
1. Sixth graders were highest. They scored significantly better than kindergarteners, first, second and third graders with a mean of 49. The fourth and fifth graders did significantly better than kindergarten through grade two only. Comparing the scores of grades 4, 5 and 6 only produced no significant differences.

2. Third, fourth and fifth graders ranked second--all three did significantly better than kindergarten, first graders and second graders. (It is of interest to note that in terms of mean scores fourth graders were higher--but just barely--than fifth graders: 47.05 to 46.98. Third graders had a mean of 45.85.)

3. Kindergarteners, first and second graders were last with no significant differences among their scores. Of the three, the first graders were highest. Mean scores were: kindergarten 41.58, 1st graders, 43.63, 2nd graders, 42.82.

To make these differences among the scores easier to visualize they have been plotted on page 73.

MEAN SCORES BY GRADE LEVEL ON PARTS 1A & 2A



Top Ten Subjects from Each Grade - Part 1A: —·—·
Part 2A: ·····

All Subjects - Part 1A: ———
Part 2A: - - - -

A closer look at the results on the individual items reveals some relevant information.

1. There were statistically significant differences among grade levels on all but four items. (This points up that most of the items in part 1A were functioning as discriminators.) One of these four items was the long i- sound. The rather high means for each grade, ranging from 3.767 to 3.950, indicate that this sound (as expected) was not extremely troublesome. It had appropriately been placed as the first item on the test. One can say that all grade levels did equally well on this item. Quite to the contrary, all grade levels did equally poorly on two items which failed to show a significant difference among any grade levels, namely item 13 testing the 1- sound and item 14 testing the initial r- sound. Mean scores ranged from 1.43 to 1.80 on item 13, and from 1.01 to 1.10 on item 14. These low scores accentuate the difficulty of the uvular trilled r- sound and the 1- sound for American speakers learning German.

2. Although generally speaking the higher grade levels exceeded the lower grade levels, the first grade was significantly better than both the fourth and sixth graders on item number 11 testing the front ch- sound. Means on this item were rather low: first graders ranked highest at 1.20 and second graders were lowest with .98.

3. Sixth grade superiority appears to be most evident in consonant clusters. For example, they are significantly better

than all grades except fourth and fifth on item 16, and surpass all grades on item 18.

4. There is a high degree of agreement between the two statistical tests the subject scores were subjected to for analysis. A copy of the computer print-outs of the analyses done on the four parts of the test (1A, 1B, 2A, 2B) and the total test scores can be found in Appendix I. The reader should note that on the Mann-Whitney-U test scores the columns are conveniently labeled "higher" and "lower" group respectively. The Fisher-Student-t scores are recorded as "group 1" or "group 2." The mean scores indicate which of the two is higher.

Part 1B: Word Stress. Attention will now be focused on part 1B of the test in which the ability to imitate word stress was evaluated. It should be kept in mind that the words in this part of the test were all monosyllabic. Furthermore, the stress pattern on each of the words was made as similar as possible in order not to detract from the primary objective of this part of the test--to evaluate imitative ability of critical sounds and sound clusters embedded in single words. Since the word stress pattern was the same in each item, there seemed to be little point in doing an item by item analysis on this part of the test. The scores achieved by all subjects were very high (in terms of possible point) and very similar. Nevertheless, an analysis of composite scores by grade level was done, the results of which appear below.

Table 7: Part 1B: Word Stress--Significant Differences Among Grade Levels on Composite Scores

Higher Groups	Lower Groups
4,5,6	K,1
6	2,3

The differences among grade levels are not so numerous as in part 1A of the test, but they confirm the superiority of the older subjects--fourth, fifth and sixth graders did significantly better than kindergarten and first graders. The sixth graders did better than the second and third. It should be kept in mind that while these differences may be statistically significant, they may have limited pedagogical implications since the mean scores achieved were really very similar ranging from 39.25 for kindergarten to a perfect 40 for the sixth graders.

Part 2A: Sounds and Sound Clusters in Phrases. In order to determine how well each item was functioning in terms of discriminating among grade levels, and in order to detect any trends in imitative ability, an item-by-item analysis comparing each grade level with every other grade level was also done on this part of the test. The results appear in Table 8. The detailed analyses can again be seen in Appendix I.

Since the number of significant differences at the .05 level and above is even larger than in part 1A of the test (155 to

Table 8: Part 2A: Sounds Embedded in Phrases: Significant Differences Among Grade Levels on Each Item

Item Number	Sound Tested	Higher Groups	Lower Groups
1	[i:]	3,4,5,6	K,1
2	[e:]	NSD	NSD
3	[o:]	NSD	NSD
4	[ɔ]	1,3,6	K
5	[a]	6	K
6	[ø:]	NSD	NSD
7	[œ]	4,5,6 5,6	K,2 1,3
8	[y:]	6	2
9	[Y]	4,5,6	K,1,2
10	[a^{U}]	1,3,4,5,6 6	K 2
11	[ç]	2,3,5,6	4
12	[X]	3,4,6 3,4,5,6 6	K 1,2 3,5
13	[1]	2,3,4,6 6 6	K 1 5
14	[R]	5 6	K K,1,3,4
15	[ʀ]	NSD	NSD
16	[pf]	5,6 4 5,6	K,1 1 2
17	[kŋ]	1,2,3,4,5,6	K
18	[ʃt]	2,3,5,6 3,5,6	K 1
19	[tsv]	5 6	1,2,3 1
20	[çts]	5	K,1

130) a cross-tabulation was again done (Table 9) to make these results more meaningful and easier to interpret.

Table 9: Part 2A: Cross Tabulation of Significant Differences Among Grade Levels on Individual Items

High Grade	Low Grade K	1	2	3	4	5	6	Totals
K	X	0	0	0	0	0	0	0
1	3	X	0	0	0	0	0	3
2	3	0	X	0	1	0	0	4
3	7	2	1	X	1	0	0	11
4	7	4	3	0	X	0	0	14
5	9	8	5	2	1	X	0	25
6	12	9	6	3	2	2	X	34
Totals	41	23	15	5	5	2	0	91

Even clearer trends immerge on this cross-tabulation than on the one for part 1A. Without exception, the increase in the number of times a given grade-level achieved a significantly higher score on individual items in part 2A runs directly parallel to the increase in grade level. By the same token the frequency of appearance as "low group" decreases as the grade level increases except for grades three and four which are equal in this respect. In other words, a comparison of the scores on the individual items in part 2A of the test (critical sounds and sound clusters embedded in short sentences) very clearly indicate that

the higher the grade level, the better was the subjects' imitative ability.

There is again a large difference in the performance of the lowest three grades (K-2) as compared to the upper four grades (3-6). Within these two groups the same general trends emerge as in part 1A, an increase in performance runs almost directly parallel to an increase in grade level.

A comparison of significant differences (.05 or higher) among grade levels on the basis of <u>total scores achieved on part 2A</u> of the test yielded the following results.

Table 10: Part 2A: Significant Differences Among Grade Levels on Composite Scores

Higher Groups	Lower Groups
1,3,4,5,6	K
3,4,5,6	1,2
6	3,4

This table shows that:

1) Kindergarteners were the lowest - every other grade except the second did significantly better.

2) First and second graders were second lowest. The first graders did significantly better than kindergarteners but the second did not.

3) Third, fourth and fifth graders were next and did equally well--they all did significantly better than kindergarteners, first graders and second graders.

4) Sixth graders were the best--they did significantly better than subjects from every other grade except the fifth.

The mean scores on part 2A have also been plotted on page 73 in order to give a visual presentation of these differences in imitative ability as measured by this instrument.

Part 2A of the test will now be analyzed in further detail by looking at the results on the individual items.

1) There were statistically significant results among grade levels on all but four items in this part of the test. This shows that most items performed as a discriminator of imitative ability among grade levels of the subjects.

2) In terms of the total number of significant differences among grade levels, consonants and consonant clusters are better discriminators than vowels.

3) Short vowels were considerably better discriminators than long vowels. This is as expected, since many American speakers of German tend to have difficulty making short vowels too long.

4) The length of the utterance in terms of total syllables does not appear to be an important factor in the items' power to discriminate among grade levels, i.e. some

two- and three- syllable utterances were as effective as four-syllable ones in this respect.

6) In terms of comparison with part 1A, which (as explained earlier) must be done with extreme caution, the following observations are of interest.

a) There are considerably more significant differences among grade levels in part 2A than in part 1A.

b) A given sound embedded in a word does not appear to have the same discriminatory ability when embedded in a phrase. (Items 1,2,13 and 14 are good examples.) Both the difference in the length of the utterances and in the difference in the phonetic environment in which the sounds are embedded could be factors contributing to this.

c) Plotting the mean scores of part 1A and part 2A together (page 73) visually demonstrates very similar results on these two parts of the test. It is of interest to note, however, that kindergarteners and first graders did slightly poorer on part 2A than on part 1A. Fourth graders were almost identical on the two parts. The rest of the grades all did better on part 2A of the test. Apparently the longer utterances on part 2A had more of an effect

on the younger subjects. The older subjects may have benefitted from the practice and experience in part 1A.

Part 2B: Sentence Stress. The last part of the test to be examined separately is part 2B, which attempted to evaluate the subjects' ability to imitate sentence stress. In doing so, several factors must be pointed out:

1) The primary purpose of part 2 of the test was--as in part 1--to test the subjects' ability to imitate critical sounds and sound clusters of German.

2) The types of intonation patterns that could be tested were severely limited by the fact that the utterances were rather short--a maximum of four syllables.

3) Emphasis was on using normal German utterances. To indiscriminately change stress patterns from one sentence to the next just to see if subjects could imitate them properly was not considered an alternative.

4) The total scores for all grade levels were very high in terms of the total possible score, just as in part 1B.

A comparison of scores among grade levels showed significant differences at the .05 level or higher (see Table 11). Although there are considerably more significant differences among grade levels here than on part 1B of the instrument (which tested word stress) and the results also favor the older subjects, the mean scores were again quite similar ranging from 37.05 (kindergarten) to 39.92 (6th graders).

Table 11: Part 2B: Sentence Stress: Significant Differences Among Grade Levels on Composite Scores

Higher Groups	Lower Groups
1,2,3,4,5,6	K
3,4,5,6	1
4,5,6	2
5,6	3
6	4

The pattern that emerged here from the sixth down to the second grade showed that no grade level was significantly better than the one immediately below it, but rather significantly better than the grade two levels below it. Thus, the sixth grade was significantly better than grades 4,3,2,1 and kindergarten (but not better than grade 5) etc. The only exception to this pattern was that the first graders exceeded kindergarteners at a significant level.

To summarize one could say that greater differences among subjects were detected on the basis of the ability to produce critical sounds and sound clusters than in their ability to imitate word and sentence stress.

Total Test. Finally, the imitative ability among grade levels was compared on the basis of scores on all four parts of the test. As in every other part of the test, all differences

significant at the .05 level and above were recorded and appear below.

Table 12: Total Test: Significant Differences Among Grade Levels

Higher Groups	Lower Groups
1,3,4,5,6	K
3,4,5,6	1,2
6	3,4,5

On the basis of the results on the total instrument used in this study, the following statements can be made concerning the imitative ability of the subjects according to their grade level.

1. Sixth graders were highest--having done significantly better than subjects from all other grades.

2. Third, fourth and fifth grades did equally well and ranked second--all outscored kindergarten, first and second graders at a significant level.

3. First and second grade rated third. First graders were slightly better, since their total scores were significantly better than those of the kindergarteners.

4. Kindergarteners were last having a significantly lower score than all other grade levels except the second.

Thus the null hypothesis posed earlier in this study has been rejected. It stated that there is no significant difference in the ability of randomly selected subjects, kindergarten through

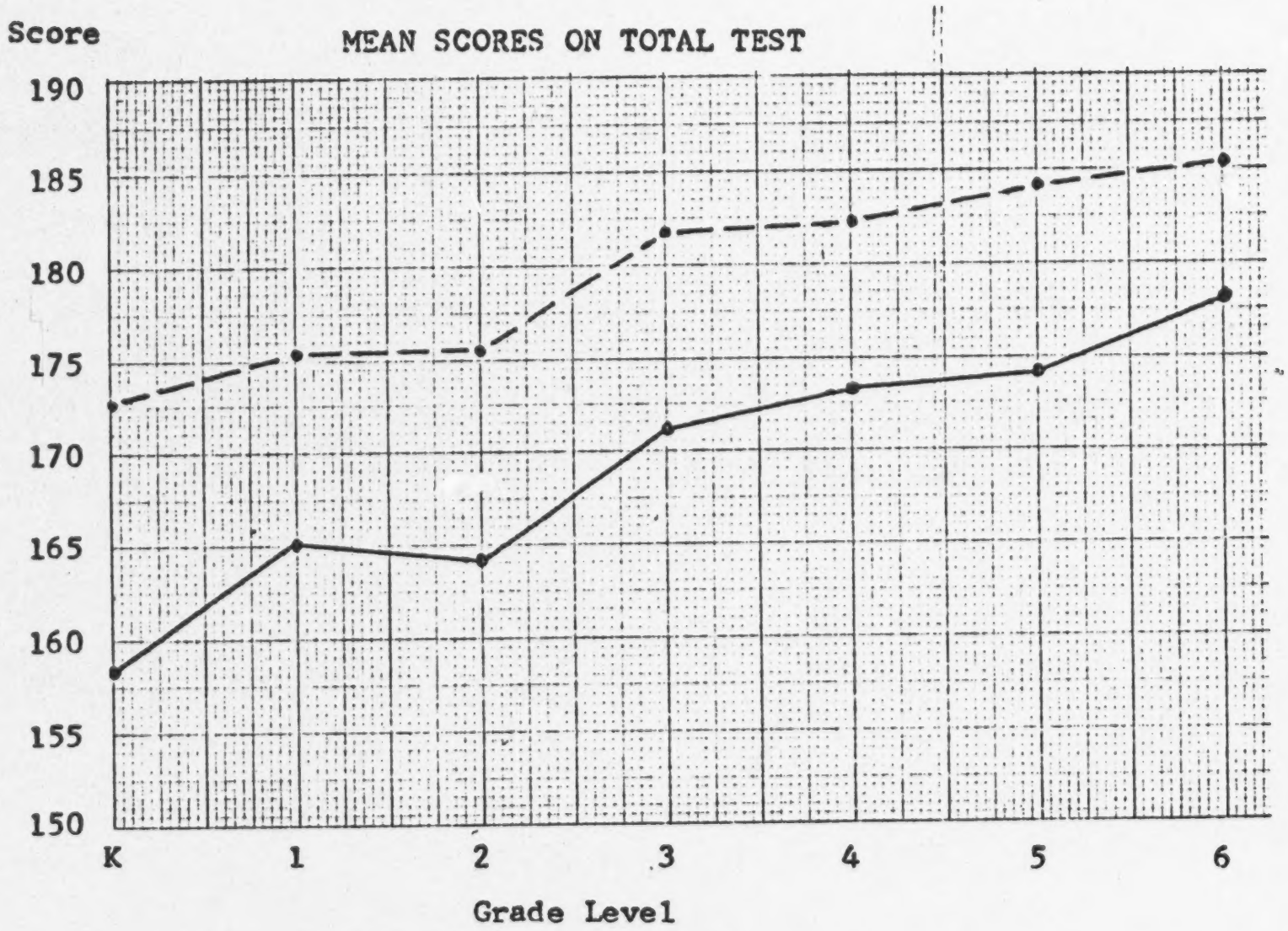
grade six to imitate critical sounds and sound clusters embedded in words and phrases, word stress, and sentence stress of German.

The plotting of the mean scores achieved on the entire test by each grade level can be seen on page 86. Both the means and standard deviations are given in Table 13.

Table 13: Rank, Mean Scores and Standard Deviations on Total Test

Grade Level	Rank	Mean	S.D.
6	1	178.20	8.25
5	2	174.08	9.27
4	3	173.42	8.87
3	4	171.33	10.64
2	6	164.15	10.78
1	5	165.05	8.75
K	7	158.31	13.19

The standard deviations indicate that there is a fairly wide range in imitative ability within each grade level. The S.D tends to decrease slightly as the grade level increases. It should be pointed out, however, that the S.D. of the first graders was second lowest--not much higher than that of the sixth grade. Furthermore, the S.D. of the fifth grade exceeded that of fourth grade. Thus there is not a direct relationship between grade level



Top Ten Subjects from Each Grade Level: - - - -

All Subjects: ————————

and standard deviations, but the rather high standard deviation of the kindergarteners is noteworthy.

Reliability. Test reliability on the entire instrument was calculated by a Pearson Product-Moment Correlation Coefficient on scores achieved by all subjects on odd and even numbered items. Adjusting this split-half correlation coefficient using the Spearman-Brown Prophecy Formula yielded a reliability coefficient of .76.

Variable 3: Sex

The sample of subjects for this study consisted of fifteen boys and fifteen girls from each grade level. Accordingly, appropriate analyses were made on the four parts of the test as well as on the results of the instrument in its entirety. These will now be discussed.

Composite scores on the four parts of the test showed that the girls achieved slightly better scores than the boys on parts 1A, 2A, and the total test. But none of these scores revealed any significant differences. In summary then, one could say that, the scores achieved by the boys and girls on any part and on all of this instrument did not reveal any significant differences in imitative ability between the two sexes.

Comparison of the Ten Best Subjects on Each Grade Level

When comparing the performance of various groups on a

given instrument, it is frequently of interest to stratify the subjects according to some criterion, and then to compare their scores again. In this case it was thought it might be valuable to make a brief comparison of the subjects with the ten highest scores from each grade level. In other words, the top third of each grade was compared. This was done on the basis of composite scores on parts 1A, 1B, 2A, 2B and total test scores. All three variables were included.

1. Judge variability: There was a significant difference only on part 1A. It should be pointed out again, that this disagreement on just one part of the test was neither to the advantage nor disadvantage of any given grade level, since the average of the scores assigned by the two judges were used in the analyses.

2. Sex: There were no significant differences between the achievement of the boys and girls on any of the four parts nor on the total test (Appendix J).

3. Grade variability: The performance of the top third of the subjects from each grade will now be compared (Appendix J).

In comparing these results (top third) with those of the total sample, the same general patterns emerged: an increase in grade level was associated with an increase in total score. Significant differences among grade levels were not as marked,

Table 14: Top Ten Subjects On Each Grade Level: Significant Differences Among Grade Levels

	Higher Groups	Lower Groups
Part 1A:	3,4,5,6 6	K,1 2
Part 1B:	NSD	NSD
Part 2A:	4,5,6 5,6	K 1,2
Part 2B:	2,3,4,5,6 4,5,6	K 1
Total Test	3,4,5,6 6 5,6	K,1 3,4 2

however. The rankings on the basis of the entire test looked like this:

1. Sixth graders were highest--they had significantly better scores than all grades except the fifth.

2. Third, fourth and fifth graders were second. There were no significant differences among their scores. Of the three, the fifth graders were highest since they were significantly better than the second graders--the third and fourth graders were not.

3. Kindergarteners, first and second graders were last. There were no significant differences among their scores.

Mean scores for parts 1A, 2A and the total test have been plotted on pages 73 and 86. This visually demonstrates the

similarity of the scoring patterns of the top third and the total population. Notable exceptions are the high scores of the top third of the third grade and that of the top third of the second grade on part 1A.

The rank, mean scores and standard deviations of the top third of the subjects from each grade level are given in Table 15.

Table 15: Mean Scores of the Top Ten Subjects in Each Grade Level

Grade	Rank	Mean	Standard Deviation
6	1	185.70	2.70
5	2	184.20	5.53
4	3	182.20	2.95
3	4	181.90	3.87
2	5	175.50	9.42
1	6	175.40	4.71
K	7	172.90	5.02

In summary it is noteworthy that:

1. There is a directly inverse relationship between the mean scores and the grade level.

2. The mean scores of the top third of the subjects from each grade level are approximately 10 points higher and the standard deviations are considerably lower than those of the total

population. Both of these results are to be expected.

3. In terms of achievement by grade, the same general trends appear as for the total population.

4. The null hypothesis stated earlier in this study was again rejected.

Ranking of Sounds and Sound Clusters by Difficulty

Since some sounds and sound clusters were clearly more difficult to imitate than others it was decided to rank them according to difficulty on the basis of the mean scores achieved. An inventory of this type could have considerable pedagogical value for any of a number of reasons: 1) Appropriate amounts of time could be spent in remedial work. 2) More effective strategies for teaching troublesome sounds and sound clusters might be explored. 3) It might have implications for the production of materials suitable for initial stages of language learning. 4) It might expedite more individualization of instruction.

Since the critical sounds and sound clusters were not in the same phonetic environment in the two parts of the test, they were ranked separately. This, of course, provided another avenue of comparing the two parts of the test. The sounds were ranked from the easiest to the most difficult according to the mean scores for all subjects (Table 16).

The data in Table 16 shows some rather interesting results:

Table 16: Rank Order of Sounds According to Difficulty

Part 1A			Part 2A		
Sound Tested	Rank	Mean	Sound Tested	Rank	Mean
[i:]	1	3.85	[i:]	1	4.03
[kn]	2	3.20	[Y]	2	3.26
[Y]	3	2.94	[œ]	3	3.22
[a]	4	2.92	[a]	4	2.87
[ɔ]	5	2.68	[kn]	5	2.76
[œ]	6	2.63	[e:]	6	2.67
[pf]	7	2.58	[1]	7.5	2.58
[x]	8	2.51	[ø:]	7.5	2.58
[o:]	9	2.40	[ɔ]	9	2.41
[ʃt]	10	2.38	[o:]	10	2.32
[e:]	11	2.35	[ʃt]	11	2.24
[ʀ]	12	2.17	[a$^{\ddot{U}}$]	12	2.18
[a^{U}]	13	2.10	[ʀ]	13	1.96
[ø:]	14	2.07	[R]	14	1.87
[tsv]	15	1.80	[tsv]	15	1.75
[1]	16	1.56	[pf]	16	1.51
[ɕts]	17	1.51	[x]	17	1.48
[y:]	18	1.48	[y:]	18	1.41
[R]	19	1.08	[ɕts]	19	1.12
[ɕ]	20	1.05	[ɕ]	20	1.05

1. Keeping in mind that the highest possible score on each item was five (albeit for a native pronunciation), the overall performance of the subjects was rather low. This is not to say that the subjects did poorly, but rather that imitating words or phrases containing critical sounds or sound clusters with a native-like pronunciation after hearing them modelled three times can be a rather demanding task.

2. Consonants and consonant clusters tended to be more difficult to imitate than vowels.

3. The performance of subjects on short vowels was consistently better than on long vowels. Exceptions were the /i:/ sound (which ranked highest on both parts of the test) and the /ɔ/ which ranked ninth in part 2A of the test.

4. The relative difficulty of the four short vowels was quite similar--including the short u- and o- umlauts. They ranked consecutively from three to six in part 1A, for example. Even the short a- sound ranked lower than these umlauts three of four times on the two parts of the test. The high rank of the /Y/ was surprising. The difference in performance on the long vowels was considerably greater. The long umlauted vowels /y:/ and /ø:/ tended to rank lower than the long non-umlauted vowel /o:/ and the /e:/. The latter two vowels are often diphthongized by American speakers of German.

5. The difference in rank of the two r- sounds indicates that the initial uvular, trilled r-sound tends to be more difficult

to produce than the final r- sound which is frequently reduced to almost a schwa by some native speakers of German.

6. The rather low ranking of the diphthong [aᵁ] was somewhat unexpected.

7. The front ch- sound and the consonant cluster containing it ([ʃçts]), ranked among the most difficult to imitate.

8. The performance on a given item was probably dictated by the difficulty of the critical sound or sound cluster involved rather than the length of the utterance. For example, of the four items ranked highest in part 2A of the test, two were three syllables in length and two were four syllables long.

9. In spite of the fact that the sounds and sound clusters were embedded in different phonetic environments in the two parts of the test, there was a considerable amount of similarity in their rank order. For example, five items had identical rankings: [i:] 1, [a] 4, [tsv] 15, [y:] 18, [ç] 20. Five more had just one position separating them: [Y] 3,2, [o:] 9,10; [ʃt] 10,11; [aᵁ] 13,12; [ʀ] 12,13. On the contrary several were separated by five or more rankings: [e:] [ɔ:] [ʀ]. Three were as far apart as nine rankings: [l], [x] and [pf]. On the first four of the latter six sounds, the higher ranking was always achieved on the second part of the test--sounds embedded in phrases. Overall, however, where there was a difference in the rank order of the items, the higher rank was in part 1A nine out of fifteen times.

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this study was to identify the relationship between subjects from various grade levels and their ability to imitate selected sound features of German. The subjects consisted of two hundred ten randomly selected children, fifteen boys and fifteen girls from each grade level, kindergarten through grade six. These subjects were asked to imitate forty German utterances arrived at on the basis of a pilot test conducted with thirty-nine subjects of the same age range. The test stimuli consisted of twenty mono-syllabic words and twenty phrases ranging in length from two to four syllables. Only one critical sound or sound cluster was embedded in each of the utterances which were modeled for prerecording by a native German linguist.

The test was administered in two of the new and ultra-modern language laboratories located in the Listening Center in the Dieter Cunz Hall of Languages on The Ohio State University Campus. Test stimuli were modeled three times. Responses of the subjects were recorded on magnetic tape and were subsequently scored by two highly qualified native-speaking judges. The production of individual sounds and sound clusters was scored

on a six-point scale and word and sentence stress were evaluated on a three-point scale.

The scores were then subjected to two statistical tests (Mann-Whitney U and Fisher-Student t) to see if the differences among the subject age groups were significant.

The principal results of the investigation showed the following:

1. The maximum length of utterances in an instrument of this type should be about four syllables.

2. The sequence of sounds (both critical and non-critical) within a given utterance seems to affect the accuracy with which subjects can imitate stimuli.

3. Test stimuli should be modeled at least three times before subjects are asked to respond.

4. Subjects of this age tend to make the same general types of pronunciation errors in a target language as children do in their native tongue. Examples are: substitutions, omissions, additions and inversions.

5. The native language is a source of interference for the subjects.

6. It is possible for two judges to evaluate the production of critical sounds and sound clusters on a five or six point scale with a high degree of agreement.

7. With a possible score of five the average achieved by all subjects on all items evaluating sounds and sound clusters

was 2.26, which indicates that imitating critical sounds and sound clusters accurately is a demanding task for second-language learners.

8. Judging on the basis of the mean scores on each item, some sounds and sound clusters are more difficult to imitate than others.

9. The scores achieved on each item depend more on the difficulty of a sound or sound cluster within an item than on the length of the utterance.

10. An analysis of three variables (judges, sex and grade level) was made. Based on scores achieved on the total instrument designed for this study, the following conclusions were reached:

a) There was no significant difference in the scores assigned to the subjects by the two judges.

b) There was no significant difference in achievement between the 105 male and 105 female subjects who participated in this study.

(c) There was a significant difference in the ability of the subjects from the seven grade levels to imitate selected sound features of German. This statement is made on the grounds that:

(1) The sixth graders ranked highest with scores significantly better than all other grade levels.

(2) The third, fourth and fifth graders did equally well and ranked second having outscored all three lower grades at a significant level.

(3) The first and second graders rated third. First graders were slightly better since they achieved a significantly better score than kindergarteners.

(4) The kindergarteners were last having a significantly lower score than subjects from all other grades except the second.

11. A comparison of the top ten subjects from each grade level produced similar results. There was no significant difference in judge variability or achievement according to sex. Although this analysis did not produce quite as marked differences among grade levels, the same general pattern appeared--imitative ability increased with grade level. More specifically the results showed:

a) The sixth graders were highest--they scored significantly better than all grade levels except the fifth.

b) The third, fourth and fifth graders were second. There were no significant differences among their scores. Of the three the fifth graders were highest since they were significantly better than the second graders--the third and fourth graders were not.

c) The kindergarteners, first and second graders were the weakest subjects and rated third. There was no

significant difference among their scores.

Although this research was conducted with subjects of a similar socio-economic level, from the same geographic area and under rather carefully specified conditions, this writer believes that the results need to be considered in terms of the following three factors: 1) generalizability, 2) implications for instructional strategies and materials, 3) suggestions for further research. These observations and recommendations are being made out of a deep concern for the future of foreign language teaching in this country, and in the hope that they might be a valuable contribution to the profession.

1. Generalizability

One of the first questions that comes to mind is, to what populations, if any, are these results generalizable? There would seem to be no reason to believe that any different results would be achieved if this study were to be repeated using the same instrument with a similar sample of subjects under similar conditions. A question of more far-reaching consequences is, of course, could one expect similar results if the critical sounds and sound clusters had been selected from a different foreign language. In other words, is a given grade level's ability (or lack of ability) to imitate the sounds of German an indication of their ability to imitate the sounds of other languages? Although only concrete research could respond to this inquiry with any degree of accuracy, speculation would lead one to believe that languages with similar

articulatory problems could conceivably produce similar outcomes. For example, an inventory of French sounds that might be included in a test of this nature might be the r- and 1- sounds, the undiphthongized [o:] and [e:], etc. Since the French variety of these sounds would not be totally unlike those of German, they might pose similar problems for subjects who had not had any exposure to the language. In summary, it might be said that there would seem to be no apparent reason to expect significantly different results provided the languages used were equally dissimilar from English.

This study would appear to have implications for the construction of instruments suitable and necessary for measuring imitative ability. The feasibility of using nonsense words and phrases might even be considered. In other words, it would be conceivable to construct an instrument that was non-language specific in an effort to test the ability of subjects to imitate selected sound features of more than one language. This was not done for several reasons, namely because of the problem of providing an authentic and natural model to be imitated, and perhaps even more important, in order to facilitate objective and accurate scoring. In this connection it should be pointed out again that several consultants voiced concern about using natural German utterances for the very reasons mentioned here.

Although no attempt was made to measure the effect of memory on imitative ability, the results on part 2 of the pilot

test seemed to imply that memory might indeed be an important factor in one's ability to imitate critical sounds and sound clusters. Decreasing the total length of the utterances by several syllables (and thus the memory task) plus an additional repetition of the cues appeared to have a very positive effect on the ability of the subjects to imitate the native German model in the final test.

2. Implications for Instruction Strategies and Materials:

a) With the ability to imitate the sounds of German in this study being inversely related to grade level, the whole issue of an optimal age for language acquisition may be raised anew. Assuming that the findings of this study can be generalized to similar populations and learning conditions and if imitative ability were the sole criterion in determining the optimal time for a random group of students age kindergarten to grade six to begin second-language study, one could establish the following priorities:

First: sixth grade,

Second: third, fourth or fifth grade,

Third: first or second grade,

Fourth: kindergarten.

If under the same conditions one had more select students, i.e. some with better imitative ability, these priorities would shift somewhat as follows:

First: sixth grade,

Second: third, fourth or fifth grades,

Third: kindergarten, first or second grades.

b) Careful consideration should be given to the length and difficulty of utterances foreign languages students are expected to imitate. It should be kept in mind that there was only one critical sound or sound cluster per utterance in the instrument constructed for this study. The focus was totally on imitating sequences of sounds. Furthermore, the utterances were short. In normal speech, utterances are frequently more than one or two syllables long. If one adds to that the burden of meaning and grammar, it should be clear that the task expected of students in our foreign language classes is not a small one.

c) Perhaps more than simple modeling of foreign language words and sounds is necessary to achieve accurate pronunciation. It might be extremely valuable to explain to students the place and manner of articulation as well as the organ of articulation.

d) Due to the fact that subjects did not score extremely well on critical sounds and sound clusters embedded in words and utterances, it may be beneficial to practice them in total isolation.

e) Since some sound clusters caused a considerable amount of difficulty, they might be broken up into their

individual components for mastery to make their pronunciation more manageable.

f) Since two competent judges did not always agree on the scores achieved by subjects, it would be wise for teachers to consult with each other frequently on topics pertaining to student performance, especially where some degree of subjective judgment is required as in the ability of students to produce critical sounds correctly.

g) More attention might be devoted to the developing and preparation of materials and teaching strategies which give attention to the problems of imitating critical sounds and sound clusters. For example, it might be wise to concentrate on a select few problems for mastery, and then proceed to others.

h) Since some learning tasks are apparently more difficult than others, more of an effort should be made to discover the difficult ones and teaching should put more emphasis on helping students with those that are repeatedly troublesome. Perhaps we could improve our materials and teaching techniques if we analyzed more closely the types of mistakes students make.

i) Carefully devised materials of a remedial nature should be developed to help students with poor imitative ability overcome their problems after they have been carefully diagnosed.

j) Judging on the basis of the benefits reaped from the pilot test in this study, all new materials and tests should be carefully explored on small groups of students before they are used on the masses.

3. Implications for Further Research:

The results of this study indicate that investigations of the following nature might be profitable.

a) Similar studies with variations such as:

(1) Critical sounds and sound clusters selected from languages other than German.

(2) Subjects of other ages, different geographic areas and various socio-economic levels.

b) Other aspects of foreign language acquisition should be explored to see if there is a difference in students' achievement relative to age.

c) The relationship of subjects' ability to discriminate among critical sounds and sound clusters and the ability to imitate them.

d) The effect of I.Q. and foreign language aptitude on imitative ability.

e) The effect of explaining place, manner and organ of articulation to subjects of various age groups.

f) An identification and classification of errors that subjects make according to grade level.

g) Memory as a factor in imitative ability.

h) Since there is a considerable degree of variance in subjects' ability to imitate critical sounds and sound clusters, an attempt should be made to identify what factors contribute most to student achievement in this area.

In conclusion this writer wishes to emphasize again, that the ability to imitate critical sound features of a foreign language is just one aspect (but an important one) of second language acquisition. Other factors such as the mastery of vocabulary, morphology and syntax must not be overlooked. Nor was it the intent of this investigation to measure the ability of subjects from various grade levels to learn how to imitate the sound features of German over a period of time in an instructional situation. Further research is also necessary to determine precisely what accounts for the significant differences among grade levels in this study. For example, the imitative task might depend on factors such as: acoustic perception, coding and memory in addition to actual articulatory functioning. To make sweeping generalizations on the basis of one study would be a mistake of serious dimensions--one which a profession treading on delicate soil can ill afford.

Appendix A: Score Sheet: Pilot Sound Production Test

SCORE SHEET: PILOT SOUND PRODUCTION TEST

Scorer: _____ Subj.#____ Grade:_____ Name:_____

Part 1: Single Words

	Sound Tested	Model word	Score (A)	Word Stress (B)
1.)	i:	die	0 1 2 3 4 5	0 1 2 3 4 5
2.)	e:	Tee	0 1 2 3 4 5	0 1 2 3 4 5
3.)	o:	wo	0 1 2 3 4 5	0 1 2 3 4 5
4.)	ɔ	Gott	0 1 2 3 4 5	0 1 2 3 4 5
5.)	a	hat	0 1 2 3 4 5	0 1 2 3 4 5
6.)	ø:	mögen	0 1 2 3 4 5	0 1 2 3 4 5
7.)	œ	gönnen	0 1 2 3 4 5	0 1 2 3 4 5
8.)	y:	müde	0 1 2 3 4 5	0 1 2 3 4 5
9.)	Y	müsse	0 1 2 3 4 5	0 1 2 3 4 5
10.)	aU	Baum	0 1 2 3 4 5	0 1 2 3 4 5
11.)	ç	ich	0 1 2 3 4 5	0 1 2 3 4 5
12.)	x	acht	0 1 2 3 4 5	0 1 2 3 4 5
13.)	l	lieb	0 1 2 3 4 5	0 1 2 3 4 5
14.)	R	Rad	0 1 2 3 4 5	0 1 2 3 4 5
15.)	ʁ	ihr	0 1 2 3 4 5	0 1 2 3 4 5
16.)	pf	Pfeife	0 1 2 3 4 5	0 1 2 3 4 5
17.)	kn	Kniff	0 1 2 3 4 5	0 1 2 3 4 5
18.)	ʃt	Stimme	0 1 2 3 4 5	0 1 2 3 4 5
19.)	tsv	Zweck	0 1 2 3 4 5	0 1 2 3 4 5
20.)	çts	Gesichts	0 1 2 3 4 5	0 1 2 3 4 5
			Score:_____	Score:_____

Part 2: Phrases

	Sound Tested	Model Utterance	Score (A)	Sentence Stress	Score (B)
1.)	i:	Sie tut das nie.	0 1 2 3 4 5	- - - ↘	0 1 2 3 4 5
2.)	e:	Sie nehmen uns mit.	0 1 2 3 4 5	- - - - ↘	0 1 2 3 4 5
3.)	o:	Im Boot sind sie.	0 1 2 3 4 5	- ↗ - →	0 1 2 3 4 5
4.)	ɔ	Die Sonne scheint.	0 1 2 3 4 5	- ↗ - →	0 1 2 3 4 5
5.)	a	Sie nimmt das Heft	0 1 2 3 4 5	- - - ↘	0 1 2 3 4 5
6.)	ø:	Schön ist sein Wagen.	0 1 2 3 4 5	↗ - - - →	0 1 2 3 4 5
7.)	œ	Kann man das öffnen?	0 1 2 3 4 5	- - - ↗ ↑	0 1 2 3 4 5
8.)	y:	Uwe muss üben.	0 1 2 3 4 5	- - - ↗ →	0 1 2 3 4 5
9.)	Y	Ist das dünn?	0 1 2 3 4 5	- - ↗↑	0 1 2 3 4 5
10.)	aU	Inge kauft etwas.	0 1 2 3 4 5	- - ↗ - →	0 1 2 3 4 5
11.)	ç	Sie haben heute Pech.	0 1 2 3 4 5	- - - - - ↘	0 1 2 3 4 5
12.)	x	Einen Kuchen, bitte!	0 1 2 3 4 5	- - ↗ - - →	0 1 2 3 4 5
13.)	l	Ist es lang?	0 1 2 3 4 5	- - ↗↑	0 1 2 3 4 5
14.)	R	Meine Mutti ruft an.	0 1 2 3 4 5	- - ↗ - - →	0 1 2 3 4 5
15.)	ʁ	Mein Neffe ißt mehr.	0 1 2 3 4 5	- - - - ↘	0 1 2 3 4 5
16.)	pf	Ein Hut mit Pfiff.	0 1 2 3 4 5	- - - ↘	0 1 2 3 4 5
17.)	kn	Wie knapp!	0 1 2 3 4 5	- ↘	0 1 2 3 4 5
18.)	ʃt	Sie stammt aus Hessen.	0 1 2 3 4 5	- - - ↗ →	0 1 2 3 4 5
19.)	tsv	Zwischen ihm und uns.	0 1 2 3 4 5	- - ↗ - →	0 1 2 3 4 5
20.)	çts	Heute nichts, danke.	0 1 2 3 4 5	- - ↖ - →	0 1 2 3 4 5
		Score:_____		Score:_____	
		TOTAL SCORE:_____		TOTAL SCORE:_____	

Appendix B: Directions for Pilot Test

Appendix B

DIRECTIONS FOR PILOT TEST: (April, 1971)

1. Thank you for being with us today. We are glad you could come. Each of you has a card with a number on it. Would you please say this number when the red light in your booth comes on? (8 seconds) Just to make sure, say it one more time when your light comes on again. (8 seconds) Thank you.

2. This is an exercise to see how well you can say the words of a language different from English. You will hear someone say a word two times. When the red light in your booth comes on, say this same word one time only. Remember that you will hear each word two times, and when the red light come on, it is your turn to say the word. Speak loudly and clearly, and try to say the words just like the voice on the recording did. Now let's practice with an English word: COLUMBUS COLUMBUS (8 seconds)

Let's try some words from a different language:

HEISS HEISS (8 seconds) NEUN NEUN (8 seconds)

Are you ready? Let's begin:

3. Part 1A of Test: See Appendix A.

4. Thank you. This is the end of part 1. We will go right on to part 2. (A pause of a few minutes was inserted here.) This part of the exercise is almost like the one you've just done. You will hear a voice say short sentences instead of just one word. When

your red light comes on, say <u>the whole sentence</u> just like the voice on the recording did. Let's practice with an English sentence:

I LIVE IN OHIO I LIVE IN OHIO (8 seconds).

Now let's try one using a different language.

DAS IST INGE DAS IST INGE (8 seconds)

Are you ready? Let's begin.

6. Part 2 of Test: See Appendix A.

7. Thank you. This is the end of the exercise.

Appendix C: Letter to Subjects for Pilot Test

Appendix C

Columbus, Ohio 43221
874 Bricker Blvd.
April 17, 1971

Dear Friend:

Have you ever wondered how good you are at saying the sounds of a language other than English? This is a chance for you to find out! You and other children from Tremont School have been selected to participate in an experiment to see which grade can do it the best. Before your parents let you accept this invitation, I'm sure they will have some questions, so let me explain a few things to them.

1. I am a Ph.D. candidate at Ohio State University and am preparing to gather data for my dissertation in which I will attempt to measure what effect, if any, maturation has on a child's ability to imitate foreign language sounds.

2. This study is being done with the approval of the Upper Arlington School Administration, and is supported with a grant from the U.S. Office of Education.

3. Your child was randomly selected from students at Tremont School. The date for this project is: Saturday, April 24.

4. The subjects will hear some short prerecorded words and phrases and will be asked to repeat them. (The responses will be recorded and evaluated by an expert panel of judges.) Subjects will then listen to a few short words and utterances to see how well they can hear the differences among them. The total session should not take more than about 15 - 20 minutes. There will be a brief pause for relaxation between the two activities.

5. Round trip transportation will be provided from the Tremont School parking lot to the new language laboratories in the Dieter Cunz Hall of Languages on the OSU campus where the recordings will be made. Young children may be accompanied by an older friend or member of the family, if you desire.

6. Children participating in the study should NOT:
   a.) have any speech defects,
   b.) have any previous exposure to a foreign language other than what most children hear on popular children's programs such as: "Sesame Street" and "Lucy's Toy Shop."

7. If you are willing to have your child participate in this study, please complete the enclosed form, detach and return it to me immediately in the stamped, self-addressed envelope. Results of the study (including your child's score) will be made available to you upon request.

8. If you have any questions, please feel free to call me at:

a) 451-9559 - home, evenings.
b) 422-9261 - OSU office, P.M.

I will be waiting to hear from you, and may I thank you in advance for your assistance.

Sincerely yours,

Joe Wipf

Joe Wipf

P.S. It may be of interest to you that the Mrs. Wipf who is teaching a 6th grade at Tremont School is my wife.

Appendix D: Response Sheet from Subjects of Pilot Test

Appendix D

TO: Joe Wipf

DATE: April, 1971.

My child, _____, has my permission to participate in your study. I have indicated several convenient time choices (where possible one in the morning and one in the afternoon) in order of preference.

_____ 9:30 A.M.*

_____ 10:30 A.M. (All on Saturday, April 24.)

_____ 1:30 P.M.

_____ 2:30 P. M.

_____ Other: a) Time: _____

b) Day: _____ (Including week days)

_____ Tel. # _____
(Signature of Parent or Guardian)

I (am, am not) interested in the results of the study.
My child (will, will not) be accompanied by someone older.

- -

PLEASE DETACH HERE AND KEEP FOR YOUR OWN RECORD

*The time indicates departure time. Unless you hear otherwise, I assume your child will be at Tremont School parking lot at the time of your first choice. We will aim to be back in one hour.

The times I selected are:

1st choice: _____

2nd choice: _____

3rd choice: _____

Joe Wipf (451-9559)
874 Bricker Blvd.
Columbus, Ohio 43221

(Saturday, April 24.)

Appendix E: Score Sheet: Final Sound Production Test

SCORE SHEET: FINAL SOUND PRODUCTION TEST

Scorer: Prof. Wängler Subject: _____
Mrs. Wängler Grade: K, 1, 2, 3, 4, 5, 6.

Part 1: Single Words

	Sound Tested	Model Word	Score (A)	Word Stress (B)
1.)	i:	die	0 1 2 3 4 5	0 1 2
2.)	e:	See	0 1 2 3 4 5	0 1 2
3.)	o:	wo	0 1 2 3 4 5	0 1 2
4.)	ɔ	Gott	0 1 2 3 4 5	0 1 2
5.)	a	hat	0 1 2 3 4 5	0 1 2
6.)	ø:	schön	0 1 2 3 4 5	0 1 2
7.)	œ	gönnt	0 1 2 3 4 5	0 1 2
8.)	y:	kühn	0 1 2 3 4 5	0 1 2
9.)	Y	dünn	0 1 2 3 4 5	0 1 2
10.)	aU	Baum	0 1 2 3 4 5	0 1 2
11.)	ç	ich	0 1 2 3 4 5	0 1 2
12.)	x	Dach	0 1 2 3 4 5	0 1 2
13.)	l	Null	0 1 2 3 4 5	0 1 2
14.)	R	Rad	0 1 2 3 4 5	0 1 2
15.)	ʀ	ihr	0 1 2 3 4 5	0 1 2
16.)	pf	Pfiff	0 1 2 3 4 5	0 1 2
17.)	kn	Kniff	0 1 2 3 4 5	0 1 2
18.)	ʃt	Stamm	0 1 2 3 4 5	0 1 2
19.)	tsv	Zweck	0 1 2 3 4 5	0 1 2
20.)	çts	nichts	0 1 2 3 4 5	0 1 2

Score: _____ Score: _____ PART 1 TOTAL: _____

Part 2: Phrases

	Sound Tested	Model Utterance	Score (A)	Sentence Stress	Score (B)
1.)	i:	Sie tat es.	0 1 2 3 4 5	´ - -↓	0 1 2
2.)	e:	Das tut weh.	0 1 2 3 4 5	- - ´↓	0 1 2
3.)	o:	Im Boot.	0 1 2 3 4 5	- ´↓	0 1 2
4.)	ɔ	In Bonn.	0 1 2 3 4 5	- ´↓	0 1 2
5.)	a	Sie kann es.	0 1 2 3 4 5	- ´ -↓	0 1 2
6.)	ø:	Eine Möve.	0 1 2 3 4 5	- - ´ -↓	0 1 2
7.)	œ	Wieso öffnen?	0 1 2 3 4 5	- - ´ -↓	0 1 2
8.)	y:	Muss sie üben?	0 1 2 3 4 5	- - ´ -↑	0 1 2
9.)	Y	Gute Nüsse.	0 1 2 3 4 5	- - ´ -↓	0 1 2
10.)	aU	Am Ausgang.	0 1 2 3 4 5	- ´ -↓	0 1 2
11.)	ç	Sie hat Pech.	0 1 2 3 4 5	- - ´↓	0 1 2
12.)	x	Einen Kuchen.	0 1 2 3 4 5	- - ´ -↓	0 1 2
13.)	l	Wie lose!	0 1 2 3 4 5	- ´ -↓	0 1 2
14.)	R	Sie rufen.	0 1 2 3 4 5	- ´ -↓	0 1 2
15.)	ʀ	Sie ißt mehr.	0 1 2 3 4 5	- - ´↓	0 1 2
16.)	pf	Ein Pfund.	0 1 2 3 4 5	- ´↓	0 1 2
17.)	kn	Wie knapp!	0 1 2 3 4 5	- ´↓	0 1 2
18.)	ʃt	So stumm!	0 1 2 3 4 5	- ´↓	0 1 2
19.)	tsv	Zwischen ihnen.	0 1 2 3 4 5	´ - - -↓	0 1 2
20.)	çts	Des Gesichts.	0 1 2 3 4 5	- - ´↓	0 1 2

Score: _____ Score: _____

Sound Score: _____ Stress Score: _____

PART 2 TOTAL: _____ TOTAL SCORE: _____

Appendix F: Directions for Final Test

Appendix F

DIRECTIONS FOR FINAL TEST

(May, 1971)

I. Speaking Exercise:

A. Thank you for being with us today. We are glad you could come. We have two activities planned for you. We are sure you will enjoy both of them. To help you do well, please relax and feel at home, because you will not even need to touch any of the buttons in your booth. Please try to remember not to move the microphone in front of your mouth.

So we don't forget who you are, could you please give us both your first and last names in a moment when the red light in your booth comes on. (10 sec.) Just to make sure, please say your full name one more time when your red light comes on again. (10 sec.) Thank you.

This is an exercise to see how well you can say the words of a language different from English. To make it more interesting, we will make it a contest to see which grade in your school can do it the best: kindergarteners, 1st, 2nd, 3rd, 4th, 5th or 6th grade. You will hear someone say a word three times. When the red light in your booth comes on, say this same word one time only. You may practice the word quietly to yourself as you listen to the voice. Remember that you will hear each word three times, and when your red light comes on, it is your turn to say the word just one time. Speak loudly and clearly, and try to say the words just exactly like the voice on the recording.

Let's practice with an English word:

RED (2 sec.) RED (2 sec.) RED (8 sec.)

Now let's try some words from a different language:

HEISS (2 sec.) HEISS (2 sec.) HEISS (8 sec.)

NEUN (2 sec.) NEUN (2 sec.) NEUN (8 sec.)

Are you ready? Let's begin:

B. Part 1 of Test: see Appendix E.

C. Thank you. The next exercise is almost like the one you have just done. You will hear a voice say a short sentence instead of just one word. Remember, you will hear the sentences three times. When your red light comes on, say the sentence one time only. Try to say it just like the voice on the recording did. Let's practice one in English:

IN COLUMBUS (2 sec.) IN COLUMBUS (2 sec.)

IN COLUMBUS (8 sec.)

Now let's try some in a different language.

DAS IST INGE (2 sec.) DAS IST INGE (2 sec.) DAS IST INGE (8 sec.)

WIE HEISST DU? (2 sec.) WIE HEISST DU? (2 Sec.)

WIE HEISST DU? (8 sec.)

Are you ready: Let's begin:

D. Part 2 of test: See Appendix E.

E. Thank you. This is the end of the exercise.

Appendix G: Vita of Personnel

Appendix G

Vita of Personnel

1. Dr. Edward D. Allen, Project Director (Professor of Foreign Language Education).

Training:

B.A., Montclair State Teachers College, Montclair, N.J., 1943
M.A., University of Wisconsin, 1948
Ph.D., The Ohio State University, 1954
French Diploma, University of Grenoble, 1950
Institut de Touraine, Summer 1959
Mexico City College, six summers

Experience:

Belleville High School, New Jersey, 1943-45, French teacher
The Ohio State University, 1945-54, Instructor of Spanish and French at the Campus Demonstration Center
The Ohio State University, 1954-58, Assistant Professor, Campus School
The Chio State University, 1958-62, Associate Professor, Department of Education
Ohio Wesleyan University, 1956, Visiting Lecturer in Spanish
Staff of NDEA Summer Language Institute, Univ. of Notre Dame, 1960
Staff of NDEA Summer Language Institute, Univ. of Maine, 1961
Director, NDEA Summer Language Institute, The Ohio State Univ., 1962-64
Director, NDEA Summer Language Institute, Lyon, France, 1965-69.

Publications:

"A New Foreign Language Course for Advanced Students in High School," Modern Language Journal, 41: 121-125, March, 1957.

"Why Not Student Exchanges at the High School Level," The French Review, 31: 136-140, December, 1957.

"How to Teach Students to Think in Spanish," Modern Language Journal, 42: 139-141, March, 1958.

"Folk Festivals," Journal of Health, Physical Education, and Recreation, 30: 31-32, February, 1959.

"The Effects of the Language Laboratory on the Development of Skills In a Foreign Language," Modern Language Journal, 44: 355-358, December, 1960.

"Foreign Language in the Elementary School," Educational Research Bulletin, 40: 85-88, April, 1961.

"A Modern Language Teacher Makes Latin Drills for the Laboratory," Classical Journal, 57: 160-163, January, 1962.

"The Language Laboratory in Learning Foreign Languages," Theory Into Practice, 1: 20-24, February, 1962.

"The Education and Re-Education of Foreign Language Teachers," Modern Language Journal, Vol. 48, May, 1964.

"Preparation of Foreign Language Teachers," The Encyclopedia of Education (to be published in 1971).

"The Teaching of Foreign Languages," The Encyclopedia Americana (to be published in 1971).

The Changing Curriculum: Modern Foreign Languages, Association for Supervision and Curriculum Development, NEA. Co-authored with Frank Otto and Leona Glenn, 1968.

2. Dr. L. O. Andrews: Consultant (Professor of Teacher Education.)

Training:

B.A., Alma College (Michigan), 1926
M.A., University of Michigan, 1930
Honorary LL.D. Capital University, 1965

Experience:

History and Physics Teacher, Albuquerque, New Mexico, 1926
Critic Teacher in Social Studies, University High School, University of Michigan
Director of Secondary Student Teaching, University of Indiana, 1936
Assistant Professor of Education, The Ohio State University, 1945
Assistant to Dean of the School of Education, 1946
Second Dean of the College of Guam, 1953-54
Coordinator of Student Field Experiences, 1948-67.

Publications:

Contributor to "The Student Teacher in the Secondary School," published by Prentice-Hall in 1953.

With Dwight Curtis, coauthor of "Guiding Your Student Teacher," Prentice-Hall, 1954.

"Student Teaching" released by the Center for Applied Research in Education, Inc., New York City, 1964.

3. Roy Carlson: Technician

Training:

Franklin University - Electronic Degree
Valparaiso University - Electronic Degree
National Radio Institute - Correspondence Courses
1st Class FCC Radio Telephone Operators License

Experience:

WVKO, Columbus, Ohio (1950-62)
WMNI, Columbus, Ohio (1962-64)
Raydata, Columbus, Ohio (1964-65)
Magnetic Service Corporation, Columbus, Ohio (1965-69)
Ohio State University: Columbus, Ohio (1969-)

4. Edith Walters Cole: Reader for directions of the tests.

Training:

B.A., Business and Speech, Otterbein College, Westerville, Ohio, 1960
M.A., Public Address, Ohio University, Athens, Ohio, 1962
Ph.D. Candidate, Public Address, Indiana University, Bloomington, 1969.

Experience:

Technical Director of Theatre, Elmhurst College, 1964-65
Instructor in Speech, Ohio Wesleyan University, 1966
Instructor in Speech, Otterbein College, 1970

5. William E. DeLorenzo: Language Laboratory Proctor (final test)

Training:

B.A., Montclair State College, New Jersey, 1959
M.A., Montclair State College, New Jersey, 1964
Ph.D., The Ohio State University, Columbus, Ohio, 1971

Experience:

High School Spanish Teacher, New Jersey, 1959-67
FLES Teacher, New Jersey, 1964-67.
Assistant Professor of Spanish, Montclair State College, 1967
Teaching Associate in Foreign Language Education, The Ohio State University, 1968-71
Assistant Professor of Spanish and Secondary Education, University of Maryland, College Park, Maryland, 1971

6. Rudolfo Garcia: Chaperone (final test)

Training:

B.S. Spanish Education, Bowling Green State University, Ohio, 1964
M.A., Spanish, Indiana University, Bloomington, 1967
Ph.D. Candidate, Foreign Language Education, The Ohio State University, 1971

Experience:

High School Foreign Language teacher in California, 1964-66
Instructor in Spanish, Williams College, Williamstown, Mass., 1967-69

7. Dr. Gilbert Jarvis: Consultant

Training:

B.A., St. Norbert College, 1963
M.A.T., Purdue University, 1966
Ph.D., Purdue University, 1970

Experience:

High School Teacher of French and English, Wisconsin, 1963-65
Graduate Teaching Assistant, Purdue University, 1965-66

Graduate Instructor, Purdue University, 1966-70
Assistant Professor of Foreign Language Education,
The Ohio State University, 1970-

Publications:

"A Behavioral Observation System for Classroom Foreign Language Skill Acquisition Activities." Modern Language Journal, 52 (1968) 335-41.

"Program Evaluation," Britannica Review of Foreign Language Education, Volume I, (with William Hatfield).

"Strategies of Instruction for Listening and Reading Skills," Britannica Review of Foreign Language Education, Volume II (1970),

"Systematic Preparation of the Multiple-Choice Listening Test." NALLD Journal, 5, ii (1970): 18-25.

"The Practice Variable: An Experiment." Foreign Language Annals 4 (1971): 401-10.

"Individualized Learning--Where Can We Risk Compromise?" Modern Language Journal. To appear in October, 1971, issue.

8. Jim Keckley: Technical Supervisor

Training:

Student of Electrical Engineering, The Ohio State University, 1960
Associate Degree in Electronics Engineering, 1963

Experience:

Supervisor, F. W. Bell Inc., Columbus, Ohio, 1963-71.
Listening Center, The Ohio State University, 1971-

9. Bill Logsdon: Recording Technician

Training:

Ohio Technical Institute-Associate Degree in Electrical Engineering
Military Schools--2nd Army Non-Commissioned Officers School
Ordinance and Signal Supply Schools

Radio Teletype Schools
Radio Schools
Communication NCO School

Experience:

GFS Supply Co., Columbus, Ohio (1967-68)
U.S. Army (1945-67)
The Ohio State University, Columbus, Ohio, 1970-

10. Gary Mann: Acting Director of the Listening Center

Training:

Ohio Technical Institute
Rochester Technical Institute
Military Schools - Associate Degree in Electrical Engineering and BSEE

Experience:

National Cash Register, Dayton, Ohio (1965-67)
North American Rockwell, Columbus, Ohio (1967)
General Dynamics, Rochester, New York (1967-68)
The Ohio State University, Columbus, Ohio, 1968-

11. Jenni Karding Moulton: Consultant for pilot and final test, scorer for pilot test. Native of Germany (born in Flensburg, brought up in Berlin).

Training:

Abitur, (Greek, Latin, French), Grunewald Gymnasium, Berlin
University of Heidelberg, Dolmetscher-Institut
English, 4 semesters

Experience:

Teaching German to Civil Affairs Officers at Yale, 1943-44
Teaching German to undergraduates at Cornell University, 1946-49
Working on German Tests (Listening Comprehension, Writing, Speaking) for ETS, Princeton, N.J., 1960-64, part-time
Recording of innumerable German tests for ETS and German textbook publishers
Scorer of speaking tests for ETS

Publications:

Spoken German with William G. Moulton, U.S. Armed Forces Institute, later New York: Holt & Co., 1943.

The German Vest Pocket Dictionary, New York: Random House, 1959.

Translation: (from Dutch to English): De Bare'e Sprekende Toradjas van Midden Celebes, for Yale University's "Human Relations Area Files."

12. Dr. Paul Pimsleur: Consultant

Training:

M.A., Teachers College, Columbia University
Ph.D., French, Columbia University

Experience:

University of California, Los Angeles, 1957-61
Director of the Listening Center, The Ohio State University Columbus, 1961-70
Professor of Foreign Language Education, State University of New York, Albany, 1970-

Publications:

"A Memory Schedule," Modern Language Journal, 1967, 51, 2:73-75.

Language Aptitude Battery. New York: Harcourt, Brace and World, 1966.

"Further Study of the Transfer of Verbal Materials Across Sense Modalities," Journal of Educational Psychology, 1964, 55, 2:96-102.

"Discrimination Training in the Teaching of French Pronunciation," Modern Language Journal, 1963, 47, 5:199-203.

"Foreign Language Learning Ability," Journal of Educational Psychology, 1962, 53, 1:15-26.

"Student Factors in Foreign Language Learning: A review of the literature," Modern Language Journal, 1962, 46, 4:160-170.

"A Study of Foreign Language Learning Ability: parts I and II," in Michael Zarechnak (ed.), Report of the Twelfth Annual Round Table Meeting on Linguistics and Language Studies, Washington, D.C.: Georgetown University Press, 1961.

"Transfer of Verbal Material Across Sense Modalities," Journal of Educational Psychology, 1961, 52, 2:104-107.

For additional relevant publications, see the bibliography.

13. Dr. H. H. Wängler: Chief Consultant and scorer and model of German Stimuli (Native of Hamburg, West Germany).

Training:

Dr. Phil. (University of Hamburg) 1949
Dr. habil. (University of Hamburg) 1957

Experience:

Research Associate and Instructor (University of Hamburg, Department of Phonetics) 1949-1953
Assistant Professor (University of Hamburg, Department of Phonetics) 1953-1958
Associate Professor (University of Hamburg, Department of Phonetics) 1958-1964
Visiting Professor at the Stanford University NDEA Summer Institute, 1959, 1960, 1961, 1962
Visiting Professor at the University of Colorado NDEA German Academic Year Institute, 1962-1963
Full Professor at the University of Colorado, 1964
Visiting Professor at the University of Scranton, NDEA Summer Institute, 1968
Visiting Professor at the University of Scranton, EPDA Summer Institute, 1969
Director, University of Colorado Sound Laboratories , 1968-

Publications:

A. Dissertations

Homogenisierungsprobleme an musikalischen Beispielen aus Südost-Neuguinea, Ph.D. Dissertation, Hamburg, 1949.

Zur Tonologie des Hausa, Habilitation monograph, Hamburg, 1957.

B. Books

Atlas deutscher Sprachlaute, Berlin, Akademie-Verlag, 1958. 2nd ed. 1961, 3rd ed. 1964, 4th ed. 1968.

Grundriß einer Phonetik des Deutschen, Marburg, N.G. Elwert-Verlag, 1960, 2nd ed. 1967.

Leitfaden der pädagogischen Stimmbehandlung, Berlin, Carl Marnold-Verlag, 1961, 2nd ed. 1966.

Kleine deutsche Aussprachelehre, Marburg, N. G. Elwert-Verlag, 1962, 2nd ed. 1968.

Rangwörterbuch hochdeutscher Umgangssprache, Marburg, N. G. Elwert-Verlag, 1964.

Zur Tonologie des Hausa, Berlin, Akademie-Verlag, 1964. (Excerpts from habilitation monograph of 1957).

Instruction in German Pronunciation, St. Paul, EMC Corp., 1963. 2nd ed. 1966.

Patterns in German Stress and Intonation, St. Paul, EMC Corp., 1966.

Contemporary German, with George A. C. Scherer, New York, McGraw-Hill, 1966 (College textbook for first year German). 2nd ed. with R. L. Kyes, 1971.

An Outline of German Phonetics (Translation of Grundriß einer Phonetik des Deutschen), St. Paul, EMC Corp., 1968.

Atlas of German Speech Sounds (Translation of Atlas deutscher Sprachlaute), Berlin, Akademie-Verlag, 1968.

Deutsch unserer Zeit, with E. M. Birkmaier and K. Anderson, New York, Holt, Rinehart and Winston, 1969. (College textbook for second year German).

Physiologische Phonetik, Marburg: N. G. Elwert Verlag, 1971.

In preparation:

Voice and Voice Disorders

General and Applied Phonetics

Grundstrukturen des Deutschen

C. Handbook Articles

"Phonetik," in Pädagogisches Lexikon, Stuttgart, Kreuz-Verlag, 1961.

"Sprecherziehung," in Pädagogisches Lexikon, Stuttgart, Kreuz-Verlag, 1961.

"Über die Beziehungen zwischen gesungenen und gesprochenen Tonhöhen," in Jahrbuch für musikalische Volks- und Völkerkunde, Berlin, Walter de Gruyter & Co., 1963.

"Alalie," in Enzyklopädisches Handbuch der Sonderpädagogik, Berlin, Carl Marhold-Verlag, 1966.

"Anarthrie, Dysarthirie," in Enzyklopädisches Handbuch der Sonderpädagogik, Berlin, Carl Marhold-Verlag, 1966.

"Aphrasie, Dysphrasie," in Enzyklopädisches Handbuch der Sonderpädagogik, Berlin, Carl Marhold-Verlag, 1966.

"Ausbildung der Sprachheillehrer," in Enzyklopädisches Handbuch der Sonderpädagogik, Berlin, Carl Marhold-Verlag, 1966.

"Barbarolalie," in Enzyklopädisches Handbuch der Sonderpädagogik, Berlin, Carl Marhold-Verlag, 1966.

"Intonation," in Enzyklopädisches Handbuch der Sonderpädagogik, Berlin, Carl Marhold-Verlag, 1967.

"Phonation (Stimme, Stimmgebung)," in Enzyklopädisches Handbuch der Sonderpädagogik, Berlin, Carl Marhold-Verlag, 1968.

"Stimmumfang," in Enzyklopädisches Handbuch der Sonderpädagogik, Berlin, Carl Marhold-Verlag, 1968.

"Sprache," in Enzyklopädisches Handbuch der Sonderpädagogik, Berlin, Carl Marhold-Verlag, 1968.

D. Articles

"Dem Verständnis moderner Musik," Der Chor, VII, 1949.

"Das Dimafon in der Frequenzprüfung," Zs. f. Phonetik, 1950, p. 142ff.

"Sprechmelodie-ein charakterologisches Hilfsmittel?," Hamburger Phonet. Beitrage, 1952, p. 46ff.

"Phonetik-eine Aufgabe des theologischen Studiums," Deutsches Pfarrerblatt, 1955, No. 3.

"Phonet. Grundlegung der Gesangskunde?," Musik im Unterricht, 1955, Heft 4.

"Zur sprachlichen Bedeutung der Tonhöhe beim Sprechen und Singen," with P. Martens, Sprachforum, 1955, p. 265ff.

"Die wissenschaftliche Phonetik," Neue Zs. f. Musik, 1956, p. 76ff.

"Über südwestafrikanische Bogenlieder I," Afrika und Übersee, 1955, p. 49ff.

"Über südwestafrikanische Bogenlieder II," Afrika und Übersee, 1956, p. 163ff.

Untersuchung der Kehlkopftöne und Stimmlippenbewegung nach elektrischer Recurrensreizung," with B. Schlosshauer and R. Timcke, Archiv für Ohren-, Nasen- und Kehlkopfheilkunde und Z. f. Hals-, Nasen- und Ohrenheilkunde, 1956, p. 491ff.(Congress proceedings).

"Singen und Sprechen in einer Tonsprache (Hausa)," Zs. f. Phonetik, 1958, p. 23ff.

"Sprache und Sprechen," Wiss. Zs. d. Ernst Moritz Arndt-Universität Greifswald, 1958-59, p. 563ff.

"Die Anwendung der Phonetik auf die Sprachwissenschaft," Wiss. Zs. d. Ernst Moritz Arndt-Universität Greifswald, 1958-59, p. 569ff.

"50 Jahre Phonet. Laboratorium in Hamburg," Orbis (Bulletin international de Documentation Linguistique), 1959, p. 529ff.

"Psychophonet. Untersuchungen uber die Stimme," Zs. f. Phonetik, (Festschrift Panconcelli-Calzia) 1959, p. 335ff.

"Von Wesen und Aufgaben der Phonetik," Germanica Wratislaviensia V, 1960, p. 145ff.

"Die Röntgenkinematographie als Hilfsmittel für die Lautforschung," Zs. f. Phonetik, 1960, p. 28ff.

"Zur Geschichte des Phonetischen Laboratoriums der Universität Hamburg," Die Sprachheilarbeit, 1960, p. 206ff.

"Stimmpflege als pädagogischer Auftrag," Sprachforum, (Festschrift Trojan), und Wiss. Zs. der Martin-Luther-Universität Halle-Wittenberg, 1961, p. 1315ff.

"Neuere Ergebnisse zur Tonologie des Hausa," Proceedings of the 4th International Congress of Phonetic Sciences, Helsinki, 1961, 1962, p. 787ff.

"Über die Funkton des weichen Gaumens beim Sprechen," Wiss. Zs. d. Martin-Luther-Universitat Halle-Wittenberg, (Gedenkschrift 1963, p. 1747ff.

"Grundfragen der Stimmpädagogik," Die Sprachheilarbeit (Festschrift von Essen) June 1963, p. 37ff.

"Über die Funktion der Töne im Hausa," Zs. f. Phonetik, Sprachwiss. u. Kommunikationsforschung, 1963, p. 231ff.

"Bemerkungen zur Methodik des Fremdsprachenunterrichts," Zs. f. Phonetik, Sprachwiss. u. Kommunikationsforschung, (Festschrift Hala) 1964, p. 347ff.

"Zur Bildungsaufgabe der Schwerhörigen- und Sprachheilschulen in der heutigen Gesellschaft (Mensch und Sprache)," Heilpädagogik, Oct. 1964, p. 466ff.

"Zur Grundfrage der modernen Linguistik," German Quarterly, 1966, p. 62ff.

"Some Remarks and Observations on the Function of the Soft Palate," The NATS Bulletin, 1968, p. 24.

14. Ilse G. Wängler: Scorer for pilot and final test. (Native of Hamburg, West Germany.)

Training:

Music Academy, Lübeck, W. Germany, 1940-45

Experience:

Violinist, Hamburg, W. Germany, 1948-51
Temporary resident, Boulder, Colorado 1962-63
Permanent resident, Boulder, Colorado 1964-

15. Ardis Wipf: Chaperone

Training:

B.A., Bethel College, North Newton, Kansas, 1960
M.Ed., University of Colorado, Boulder, 1964

Experience:

Seven years elementary school teaching in Colorado, Kansas, South Dakota and Ohio.

Appendix H: Explanation of Computer Print-out Labels

Appendix H

Explanation of Computer Print-Out Labels

Variable Names:

a) Grade: Numbers correspond to grade levels except that 0.0 designates kindergarten.

b) Score: S1A = Part 1A of the test.

S2A = Part 2A of the test.

Total = Total test.

S2B = Part 2B of the test.

S1B = Part 1B of the test.

Appendix I: Data Analysis (Computer Print-outs): All Subjects

```
$
TOTALS AVERAGED OVER JUDGES
$
SIZE=7,210
     7 VARIABLES  210 OBSERVATIONS
OP,NNU,FST
DATA ON UNIT 4
GROUP=1,2
SCORE=3,4,5,6,7
VARIABLE NAMES
1 = GRADE
2 = SEX
3 = S1A
4 = S2A
5 = TOTAL
6 = S2B
7 = S1B
(T4,2F2.0,5F7.1)
```

210 OBSERVATIONS WERE COUNTED

MANN-WHITNEY U FOR GROUP VARIABLE = GRADE SCORE VARIABLE = S1A
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	30	0.0	30	353.00	818.00	0.0756	0.1511APPROX
2.00	30	0.0	30	403.00	868.00	0.2434	0.4869APPROX
3.00	30	0.0	30	281.00	746.00	0.0062	0.0124APPROX
4.00	30	0.0	30	223.00	688.00	0.0004	0.0008APPROX
5.00	30	0.0	30	238.50	703.50	0.0009	0.0018APPROX
6.00	30	0.0	30	171.50	636.50	0.0000	0.0000APPROX
1.00	30	2.00	30	392.00	857.00	0.1952	0.3904APPROX
3.00	30	1.00	30	319.50	784.50	0.0267	0.0533APPROX
4.00	30	1.00	30	249.00	714.00	0.0015	0.0029APPROX
5.00	30	1.00	30	278.00	743.00	0.0055	0.0109APPROX
6.00	30	1.00	30	190.00	655.00	0.0001	0.0001APPROX
3.00	30	2.00	30	291.00	756.00	0.0093	0.0187APPROX
4.00	30	2.00	30	236.00	701.00	0.0008	0.0015APPROX
5.00	30	2.00	30	254.50	719.50	0.0019	0.0038APPROX
6.00	30	2.00	30	187.00	652.00	0.0001	0.0001APPROX
4.00	30	3.00	30	382.00	847.00	0.1572	0.3143APPROX
5.00	30	3.00	30	392.50	857.50	0.1974	0.3949APPROX
6.00	30	3.00	30	307.50	772.50	0.0175	0.0349APPROX
4.00	30	5.00	30	427.00	892.00	0.3668	0.7336APPROX
6.00	30	4.00	30	341.50	806.50	0.0540	0.1081APPROX
6.00	30	5.00	30	319.00	784.00	0.0262	0.0524APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = GRADE SCORE VARIABLE = S2A
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	30	0.0	30	329.50	794.50	0.0373	0.0745APPROX
2.00	30	0.0	30	355.00	820.00	0.0798	0.1596APPROX
3.00	30	0.0	30	198.50	663.50	0.0001	0.0002APPROX
4.00	30	0.0	30	194.00	659.00	0.0001	0.0002APPROX
5.00	30	0.0	30	200.00	665.00	0.0001	0.0002APPROX
6.00	30	0.0	30	122.00	587.00	0.0000	0.0000APPROX
1.00	30	2.00	30	424.50	889.50	0.3529	0.7057APPROX
3.00	30	1.00	30	285.00	750.00	0.0073	0.0146APPROX
4.00	30	1.00	30	255.00	720.00	0.0020	0.0039APPROX
5.00	30	1.00	30	260.50	725.50	0.0025	0.0050APPROX
6.00	30	1.00	30	163.50	628.50	0.0000	0.0000APPROX
3.00	30	2.00	30	265.50	730.50	0.0032	0.0063APPROX
4.00	30	2.00	30	248.50	713.50	0.0014	0.0029APPROX
5.00	30	2.00	30	251.50	716.50	0.0017	0.0033APPROX
6.00	30	2.00	30	171.50	636.50	0.0000	0.0000APPROX
4.00	30	3.00	30	447.00	912.00	0.4823	0.9646APPROX
5.00	30	3.00	30	441.00	906.00	0.4470	0.8940APPROX
6.00	30	3.00	30	299.00	764.00	0.0127	0.0254APPROX
5.00	30	4.00	30	435.00	900.00	0.4122	0.8243APPROX
6.00	30	4.00	30	321.50	786.50	0.0286	0.0572APPROX
6.00	30	5.00	30	357.50	822.50	0.0855	0.1710APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = GRADE SCORE VARIABLE = TOTAL
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	30	0.0	30	307.00	768.00	0.0149	0.0297APPROX
2.00	30	0.0	30	345.00	810.00	0.0602	0.1204APPROX
3.00	30	0.0	30	191.00	656.00	0.0001	0.0001APPROX
4.00	30	0.0	30	151.00	616.00	0.0000	0.0000APPROX
5.00	30	0.0	30	142.50	607.50	0.0000	0.0000APPROX
6.00	30	0.0	30	89.00	554.00	0.0000	0.0000APPROX
1.00	30	2.00	30	419.50	884.50	0.3260	0.6519APPROX
3.00	30	1.00	30	255.00	720.00	0.0020	0.0039APPROX
4.00	30	1.00	30	221.00	686.00	0.0004	0.0007APPROX
5.00	30	1.00	30	219.00	684.00	0.0003	0.0006APPROX
6.00	30	1.00	30	128.00	593.00	0.0000	0.0000APPROX
3.00	30	2.00	30	238.50	703.50	0.0009	0.0018APPROX
4.00	30	2.00	30	204.50	669.50	0.0001	0.0003APPROX
5.00	30	2.00	30	202.00	667.00	0.0001	0.0002APPROX
6.00	30	2.00	30	150.50	615.50	0.0000	0.0000APPROX
4.00	30	3.00	30	400.50	865.50	0.2320	0.4640APPROX
5.00	30	3.00	30	401.00	866.00	0.2343	0.4686APPROX
6.00	30	3.00	30	261.00	726.00	0.0026	0.0052APPROX
4.00	30	5.00	30	444.50	909.50	0.4676	0.9352APPROX
6.00	30	4.00	30	297.00	762.00	0.0118	0.0236APPROX
6.00	30	5.00	30	307.50	772.50	0.0175	0.0351APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = GRADE SCORE VARIABLE = S28
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	30	0.0	30	269.50	734.50	0.0034	0.0067APPROX
2.00	30	0.0	30	234.00	719.00	0.0015	0.0030APPROX
3.00	30	0.0	30	170.00	635.00	0.0000	0.0000APPROX
4.00	30	0.0	30	137.00	602.00	0.0000	0.0000APPROX
5.00	30	0.0	30	86.00	551.00	0.0	0.0 APPROX
6.00	30	0.0	30	75.00	540.00	0.0	0.0 APPROX
2.00	30	1.00	30	404.00	869.00	0.2362	0.4723APPROX
3.00	30	1.00	30	336.00	801.00	0.0355	0.0711APPROX
4.00	30	1.00	30	288.00	753.00	0.0040	0.0080APPROX
5.00	30	1.00	30	224.00	689.00	0.0001	0.0001APPROX
6.00	30	1.00	30	209.50	674.50	0.0000	0.0000APPROX
3.00	30	2.00	30	394.00	859.00	0.1797	0.3594APPROX
4.00	30	2.00	30	346.00	811.00	0.0374	0.0748APPROX
5.00	30	2.00	30	282.00	747.00	0.0011	0.0022APPROX
6.00	30	2.00	30	270.00	735.00	0.0004	0.0008APPROX
4.00	30	3.00	30	394.00	859.00	0.1616	0.3231APPROX
5.00	30	3.00	30	324.00	789.00	0.0083	0.0165APPROX
6.00	30	3.00	30	307.50	772.50	0.0028	0.0056APPROX
5.00	30	4.00	30	384.00	849.00	0.0809	0.1618APPROX
6.00	30	4.00	30	368.50	833.50	0.0370	0.0740APPROX
6.00	30	5.00	30	433.00	898.00	0.3258	0.6516APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = GRADE SCORE VARIABLE = S18
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
0.0	30	1.00	30	436.00	901.00	0.3970	0.7940APPROX
2.00	30	0.0	30	390.50	855.50	0.1038	0.2075APPROX
3.00	30	0.0	30	406.00	871.00	0.1830	0.3660APPROX
4.00	30	0.0	30	356.50	821.50	0.0166	0.0331APPROX
5.00	30	0.0	30	342.00	807.00	0.0051	0.0101APPROX
6.00	30	0.0	30	330.00	795.00	0.0013	0.0027APPROX
2.00	30	1.00	30	378.00	843.00	0.0692	0.1384APPROX
3.00	30	1.00	30	392.50	857.50	0.1247	0.2494APPROX
4.00	30	1.00	30	340.00	805.00	0.0079	0.0158APPROX
5.00	30	1.00	30	325.50	790.50	0.0023	0.0045APPROX
6.00	30	1.00	30	315.00	780.00	0.0006	0.0013APPROX
2.00	30	3.00	30	435.00	900.00	0.3603	0.7206APPROX
4.00	30	2.00	30	417.50	882.50	0.1779	0.3558APPROX
5.00	30	2.00	30	403.00	868.00	0.0735	0.1471APPROX
6.00	30	2.00	30	390.00	855.00	0.0201	0.0402APPROX
4.00	30	3.00	30	402.00	867.00	0.1013	0.2026APPROX
5.00	30	3.00	30	387.50	852.50	0.0379	0.0758APPROX
6.00	30	3.00	30	375.00	840.00	0.0103	0.0206APPROX
5.00	30	4.00	30	434.50	899.50	0.2720	0.5440APPROX
6.00	30	4.00	30	420.00	885.00	0.0769	0.1538APPROX
6.00	30	5.00	30	435.00	900.00	0.1587	0.3173APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = SEX SCORE VARIABLE = S1A
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
2.00	105	1.00	105	5264.50	10829.50	0.2865	0.5731APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = SEX SCORE VARIABLE = S2A
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
2.00	105	1.00	105	5062.50	10627.50	0.1533	0.3066APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = SEX SCORE VARIABLE = TOTAL
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
2.00	105	1.00	105	5272.50	10837.50	0.2928	0.5856APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = SEX SCORE VARIABLE = S2B
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	105	2.00	105	4916.50	10481.50	0.0635	0.1270APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = SEX SCORE VARIABLE = S1B
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	105	2.00	105	5243.50	10808.50	0.1540	0.3080APPROX

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = GRADE SCORE VARIABLE = S1A

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
0.0	30.	41.583	6.197	1.00	30.	43.633	4.054	-1.4908	0.1414	2.3370	0.01
0.0	30.	41.583	6.197	2.00	30.	42.817	5.583	-0.7962	0.4291	1.2318	0.29
0.0	30.	41.583	6.197	3.00	30.	45.850	6.105	-2.6413	0.0106	1.0304	0.47
0.0	30.	41.583	6.197	4.00	30.	47.050	4.903	-3.7255	0.0004	1.5974	0.11
0.0	30.	41.583	6.197	5.00	30.	46.983	5.652	-3.4672	0.0010	1.2022	0.31
0.0	30.	41.583	6.197	6.00	30.	49.000	5.204	-4.9356	0.0000	1.4179	0.18
1.00	30.	43.633	4.054	2.00	30.	42.817	5.583	0.6374	0.5264	0.5271	0.95
1.00	30.	43.633	4.054	3.00	30.	45.850	6.105	-1.6289	0.1087	0.4409	0.98
1.00	30.	43.633	4.054	4.00	30.	47.050	4.903	-2.8922	0.0054	0.6836	0.84
1.00	30.	43.633	4.054	5.00	30.	46.983	5.652	-2.5938	0.0120	0.5144	0.96
1.00	30.	43.633	4.054	6.00	30.	49.000	5.204	-4.3811	0.0001	0.6067	0.91
2.00	30.	42.817	5.583	3.00	30.	45.850	6.105	-1.9745	0.0531	0.8365	0.68
2.00	30.	42.817	5.583	4.00	30.	47.050	4.903	-3.0680	0.0033	1.2968	0.24
2.00	30.	42.817	5.583	5.00	30.	46.983	5.652	-2.8243	0.0065	0.9760	0.53
2.00	30.	42.817	5.583	6.00	30.	49.000	5.204	-4.3626	0.0001	1.1511	0.35
3.00	30.	45.850	6.105	4.00	30.	47.050	4.903	-0.8253	0.4126	1.5503	0.12
3.00	30.	45.850	6.105	5.00	30.	46.983	5.652	-0.7336	0.4661	1.1668	0.34
3.00	30.	45.850	6.105	6.00	30.	49.000	5.204	-2.1146	0.0388	1.3761	0.20
4.00	30.	47.050	4.903	5.00	30.	46.983	5.652	0.0480	0.9619	0.7526	0.78
4.00	30.	47.050	4.903	6.00	30.	49.000	5.204	-1.4687	0.1473	0.8876	0.62
5.00	30.	46.983	5.652	6.00	30.	49.000	5.204	-1.4136	0.1628	1.1794	0.33

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = GRADE SCORE VARIABLE = S2A

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
0.0	30.	40.433	6.483	1.00	30.	43.433	4.549	-2.0399	0.0459	2.0308	0.03
0.0	30.	40.433	6.483	2.00	30.	43.083	5.340	-1.6990	0.0947	1.4737	0.15
0.0	30.	40.433	6.483	3.00	30.	46.583	4.865	-4.0860	0.0001	1.7757	0.06
0.0	30.	40.433	6.483	4.00	30.	46.883	4.885	-4.2790	0.0001	1.7613	0.07
0.0	30.	40.433	6.483	5.00	30.	47.250	5.708	-4.2498	0.0001	1.2900	0.25
0.0	30.	40.433	6.483	6.00	30.	49.283	4.305	-6.1243	0.0000	2.2683	0.02
1.00	30.	43.433	4.549	2.00	30.	43.083	5.340	0.2687	0.7891	0.7257	0.80
1.00	30.	43.433	4.549	3.00	30.	46.583	4.865	-2.5468	0.0135	0.8744	0.64
1.00	30.	43.433	4.549	4.00	30.	46.883	4.885	-2.7833	0.0073	0.8673	0.65
1.00	30.	43.433	4.549	5.00	30.	47.250	5.708	-2.8159	0.0066	0.6352	0.89
1.00	30.	43.433	4.549	6.00	30.	49.283	4.305	-5.0301	0.0000	1.1169	0.38
2.00	30.	43.083	5.340	3.00	30.	46.583	4.865	-2.6091	0.0115	1.2049	0.31
2.00	30.	43.083	5.340	4.00	30.	46.883	4.885	-2.8275	0.0064	1.1951	0.32
2.00	30.	43.083	5.340	5.00	30.	47.250	5.708	-2.8706	0.0057	0.8753	0.64
2.00	30.	43.083	5.340	6.00	30.	49.283	4.305	-4.8677	0.0000	1.5391	0.13
3.00	30.	46.583	4.865	4.00	30.	46.883	4.885	-0.2343	0.8156	0.9919	0.51
3.00	30.	46.583	4.865	5.00	30.	47.250	5.708	-0.4787	0.6340	0.7264	0.80
3.00	30.	46.583	4.865	6.00	30.	49.283	4.305	-2.2383	0.0291	1.2774	0.26
4.00	30.	46.883	4.885	5.00	30.	47.250	5.708	-0.2628	0.7936	0.7324	0.80
4.00	30.	46.883	4.885	6.00	30.	49.283	4.305	-1.9850	0.0519	1.2878	0.25
5.00	30.	47.250	5.708	6.00	30.	49.283	4.305	-1.5316	0.1311	1.7584	0.07

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = GRADE SCORE VARIABLE = TOTAL

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
0.0	30.	158.317	13.190	1.00	30.	165.050	8.753	-2.2906	0.0256	2.2706	0.02
0.0	30.	158.317	13.190	2.00	30.	164.150	10.777	-1.8443	0.0702	1.4980	0.14
0.0	30.	158.317	13.190	3.00	30.	171.333	10.638	-4.1367	0.0001	1.5372	0.13
0.0	30.	158.317	13.190	4.00	30.	173.417	8.870	-5.1159	0.0000	2.2112	0.02
0.0	30.	158.317	13.190	5.00	30.	174.083	9.269	-5.2669	0.0000	2.0249	0.03
0.0	30.	158.317	13.190	6.00	30.	178.200	8.247	-6.8833	0.0000	2.5578	0.01
1.00	30.	165.050	8.753	2.00	30.	164.150	10.777	0.3491	0.7283	0.6597	0.87
1.00	30.	165.050	8.753	3.00	30.	171.333	10.638	-2.4561	0.0171	0.6770	0.85
1.00	30.	165.050	8.753	4.00	30.	173.417	8.870	-3.6156	0.0006	0.9738	0.53
1.00	30.	165.050	8.753	5.00	30.	174.083	9.269	-3.8157	0.0003	0.8918	0.62
1.00	30.	165.050	8.753	6.00	30.	178.200	8.247	-5.8884	0.0000	1.1265	0.38
2.00	30.	164.150	10.777	3.00	30.	171.333	10.638	-2.5546	0.0133	1.0262	0.47
2.00	30.	164.150	10.777	4.00	30.	173.417	8.870	-3.5753	0.0007	1.4761	0.15
2.00	30.	164.150	10.777	5.00	30.	174.083	9.269	-3.7633	0.0004	1.3517	0.21
2.00	30.	164.150	10.777	6.00	30.	178.200	8.247	-5.5756	0.0000	1.7075	0.08
3.00	30.	171.333	10.638	4.00	30.	173.417	8.870	-0.8100	0.4213	1.4385	0.17
3.00	30.	171.333	10.638	5.00	30.	174.083	9.269	-1.0496	0.2983	1.3173	0.23
3.00	30.	171.333	10.638	6.00	30.	178.200	8.247	-2.7471	0.0080	1.6639	0.09

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
4.00	30.	173.417	8.870	5.00	30.	174.083	9.269	-0.2798	0.7806	0.9157	0.59
4.00	30.	173.417	8.870	6.00	30.	178.200	8.247	-2.1268	0.0377	1.1567	0.35
5.00	30.	174.083	9.269	6.00	30.	178.200	8.247	-1.7868	0.0792	1.2632	0.27

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = GRADE SCORE VARIABLE = S2B

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
0.0	30.	37.050	2.675	1.00	30.	38.650	1.649	-2.7418	0.0081	2.6314	0.01
0.0	30.	37.050	2.675	2.00	30.	38.583	2.391	-2.3013	0.0250	1.2514	0.27
0.0	30.	37.050	2.675	3.00	30.	39.367	1.008	-4.3642	0.0001	7.0460	0.00
0.0	30.	37.050	2.675	4.00	30.	39.567	0.834	-4.8367	0.0000	10.2881	0.00
0.0	30.	37.050	2.675	5.00	30.	39.867	0.340	-5.6249	0.0000	61.8908	0.00
1.00	30.	38.650	1.649	6.00	30.	39.917	0.261	-5.7435	0.0000	104.6838	0.00
1.00	30.	38.650	1.649	2.00	30.	38.583	2.391	0.1236	0.9021	0.4756	0.98
1.00	30.	38.650	1.649	3.00	30.	39.367	1.008	-1.9969	0.0505	2.6777	0.00
1.00	30.	38.650	1.649	4.00	30.	39.567	0.834	-2.6712	0.0098	3.9098	0.00
1.00	30.	38.650	1.649	5.00	30.	39.867	0.340	-3.8912	0.0003	23.5203	0.00
1.00	30.	38.650	1.649	6.00	30.	39.917	0.261	-4.0853	0.0001	39.7829	0.00
2.00	30.	38.583	2.391	3.00	30.	39.367	1.008	-1.6256	0.1095	5.6304	0.00
2.00	30.	38.583	2.391	4.00	30.	39.567	0.834	-2.0909	0.0409	8.2211	0.00
2.00	30.	38.583	2.391	5.00	30.	39.867	0.340	-2.8612	0.0059	49.4561	0.00
2.00	30.	38.583	2.391	6.00	30.	39.917	0.261	-2.9848	0.0041	83.6514	0.00
3.00	30.	39.367	1.008	4.00	30.	39.567	0.834	-0.8234	0.4137	1.4601	0.16
3.00	30.	39.367	1.008	5.00	30.	39.867	0.340	-2.5316	0.0141	8.7838	0.00
3.00	30.	39.367	1.008	6.00	30.	39.917	0.261	-2.8448	0.0061	14.8571	0.00
4.00	30.	39.567	0.834	5.00	30.	39.867	0.340	-1.7937	0.0781	6.0158	0.00
4.00	30.	39.567	0.834	6.00	30.	39.917	0.261	-2.1564	0.0352	10.1752	0.00
5.00	30.	39.867	0.340	6.00	30.	39.917	0.261	-0.6278	0.5326	1.6914	0.08

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = GRADE SCORE VARIABLE = S1B

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
0.0	30.	39.250	1.741	1.00	30.	39.333	1.098	-0.2181	0.8282	2.5129	0.01
0.0	30.	39.250	1.741	2.00	30.	39.667	0.907	-1.1433	0.2576	3.6844	0.00
0.0	30.	39.250	1.741	3.00	30.	39.533	1.231	-0.7157	0.4771	1.9989	0.03
0.0	30.	39.250	1.741	4.00	30.	39.917	0.367	-2.0181	0.0482	22.4368	0.00
0.0	30.	39.250	1.741	5.00	30.	39.983	0.091	-2.2657	0.0272	363.5464	0.00
0.0	30.	39.250	1.741	6.00	30.	40.000	0.0	-2.3204	0.0239	0.0	1.00
1.00	30.	39.333	1.098	2.00	30.	39.667	0.907	-1.2605	0.2125	1.4662	0.15
1.00	30.	39.333	1.098	3.00	30.	39.533	1.231	-0.6529	0.5164	0.7954	0.73
1.00	30.	39.333	1.098	4.00	30.	39.917	0.367	-2.7131	0.0088	8.9286	0.00
1.00	30.	39.333	1.098	5.00	30.	39.983	0.091	-3.1770	0.0024	144.6719	0.00
1.00	30.	39.333	1.098	6.00	30.	40.000	0.0	-3.2697	0.0018	0.0	1.00
2.00	30.	39.667	0.907	3.00	30.	39.533	1.231	0.4696	0.6404	0.5425	0.95
2.00	30.	39.667	0.907	4.00	30.	39.917	0.367	-1.3760	0.1741	6.0897	0.00

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
2.00	30.	39.667	0.907	5.00	30.	39.983	0.091	-1.8711	0.0664	98.6719	0.00
2.00	30.	39.667	0.907	6.00	30.	40.000	0.0	-1.9796	0.0525	0.0	1.00
3.00	30.	39.533	1.231	4.00	30.	39.917	0.367	-1.6067	0.1135	11.2247	0.00
3.00	30.	39.533	1.231	5.00	30.	39.983	0.091	-1.9630	0.0544	181.8749	0.00
3.00	30.	39.533	1.231	6.00	30.	40.000	0.0	-2.0413	0.0458	0.0	1.00
4.00	30.	39.917	0.367	5.00	30.	39.983	0.091	-0.9482	0.3470	16.2031	0.00
4.00	30.	39.917	0.367	6.00	30.	40.000	0.0	-1.2214	0.2269	0.0	1.00
5.00	30.	39.983	0.091	6.00	30.	40.000	0.0	-0.9839	0.3293	0.0	1.00

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = SEX SCORE VARIABLE = S1A

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
1.00	105.	44.986	6.497	2.00	105.	45.562	5.363	-0.6975	0.4863	1.4673	0.03

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = SEX SCORE VARIABLE = S2A

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
1.00	105.	44.871	5.643	2.00	105.	45.686	0.168	-0.9933	0.3217	0.8378	0.82

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = SEX SCORE VARIABLE = TOTAL

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
1.00	105.	168.814	12.124	2.00	105.	169.629	11.769	-0.4915	0.6236	1.0612	0.38

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = SEX SCORE VARIABLE = S2B

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
1.00	105.	39.229	1.578	2.00	105.	38.771	2.039	1.8080	0.0721	0.5991	1.00

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = SEX SCORE VARIABLE = S18

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
1.00	105.	39.729	0.931	2.00	105.	39.610	1.100	0.8425	0.4005	0.7170	0.95

Appendix J: Data Analysis (Computer Print-outs): Top 10 Subjects from Each Grade Level

```
$
TOTALS AVERAGED OVER JUDGES
$
SIZE=7,70
     7 VARIABLES    70 OBSERVATIONS
OP,PWU,FST
DATA ON UNIT 4
GROUP=1,2
SCORE=3,4,5,6,7
VARIABLE NAMES
1 = GRADE
2 = SEX
3 = S1A
4 = S2A
5 = TOTAL
6 = S2B
7 = S1B
(T4,2F2.0,5F7.1)

        70 OBSERVATIONS WERE COUNTED

MANN-WHITNEY U FOR GROUP VARIABLE =   GRADE     SCORE VARIABLE =   S1A
     WITH PROBABILITY OF THIS OR SMALLER U
  HIGHER GROUP   N1    LOWER GROUP   N2        U      RANK SUM  PROB (1-TAIL)  (2-TAIL)
      1.00       10       0.0       10     46.00     101.00      0.3807     0.7615APPROX
      2.00       10       0.0       10     41.00      96.00      0.2474     0.4948APPROX
      3.00       10       0.0       10     21.00      76.00      0.0137     0.0274APPROX
      4.00       10       0.0       10     24.50      79.50      0.0265     0.0529APPROX
      5.00       10       0.0       10     25.00      80.00      0.0291     0.0581APPROX
      6.00       10       0.0       10     13.50      68.50      0.0028     0.0056APPROX
      2.00       10       1.00      10     49.50     104.50      0.4849     0.9697APPROX
      3.00       10       1.00      10     16.00      71.00      0.0050     0.0100APPROX
      4.00       10       1.00      10     19.00      74.00      0.0095     0.0189APPROX
      5.00       10       1.00      10     23.00      78.00      0.0204     0.0408APPROX
      6.00       10       1.00      10      8.00      63.00      0.0007     0.0015APPROX
      3.00       10       2.00      10     29.00      84.00      0.0560     0.1120APPROX
      4.00       10       2.00      10     30.00      85.00      0.0651     0.1301APPROX
      5.00       10       2.00      10     28.00      83.00      0.0477     0.0954APPROX
      6.00       10       2.00      10     24.00      79.00      0.0245     0.0490APPROX
      3.00       10       4.00      10     40.00      95.00      0.2238     0.4476APPROX
      3.00       10       5.00      10     46.50     101.50      0.3952     0.7903APPROX
      6.00       10       3.00      10     42.50      97.50      0.2844     0.5687APPROX
      5.00       10       4.00      10     48.50     103.50      0.4547     0.9093APPROX
      6.00       10       4.00      10     29.50      84.50      0.0598     0.1195APPROX
      6.00       10       5.00      10     38.00      93.00      0.1817     0.3634APPROX
```

MANN-WHITNEY U FOR GROUP VARIABLE = GRADE SCORE VARIABLE = S2A

WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	10	0.0	10	35.00	90.00	0.1278	0.2556APPROX
2.00	10	0.0	10	47.00	102.00	0.4101	0.8201APPROX
3.00	10	0.0	10	29.00	84.00	0.0559	0.1119APPROX
4.00	10	0.0	10	23.00	78.00	0.0204	0.0407APPROX
5.00	10	0.0	10	14.00	69.00	0.0032	0.0063APPROX
6.00	10	0.0	10	14.00	69.00	0.0032	0.0064APPROX
1.00	10	2.00	10	43.50	98.50	0.3114	0.6228APPROX
3.00	10	1.00	10	43.00	98.00	0.2977	0.5955APPROX
4.00	10	1.00	10	27.00	82.00	0.0408	0.0816APPROX
5.00	10	1.00	10	21.00	76.00	0.0139	0.0279APPROX
6.00	10	1.00	10	15.50	70.50	0.0044	0.0088APPROX
3.00	10	2.00	10	35.00	90.00	0.1281	0.2561APPROX
4.00	10	2.00	10	30.00	85.00	0.0650	0.1300APPROX
5.00	10	2.00	10	25.00	80.00	0.0291	0.0582APPROX
6.00	10	2.00	10	19.50	74.50	0.0105	0.0209APPROX
4.00	10	3.00	10	41.00	96.00	0.2478	0.4956APPROX
5.00	10	3.00	10	33.00	88.00	0.0990	0.1979APPROX
6.00	10	3.00	10	28.00	83.00	0.0478	0.0957APPROX
5.00	10	4.00	10	42.00	97.00	0.2715	0.5429APPROX
6.00	10	4.00	10	33.50	88.50	0.1056	0.2113APPROX
6.00	10	5.00	10	42.00	97.00	0.2720	0.5440APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = GRADE SCORE VARIABLE = TOTAL

WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	10	0.0	10	32.00	87.00	0.0863	0.1726APPROX
2.00	10	0.0	10	48.00	103.00	0.4397	0.8794APPROX
3.00	10	0.0	10	7.00	62.00	0.0006	0.0011APPROX
4.00	10	0.0	10	4.50	59.50	0.0003	0.0006APPROX
5.00	10	0.0	10	5.50	60.50	0.0004	0.0008APPROX
6.00	10	0.0	10	1.00	56.00	0.0001	0.0002APPROX
1.00	10	2.00	10	43.50	98.50	0.3111	0.6221APPROX
3.00	10	1.00	10	15.50	70.50	0.0045	0.0089APPROX
4.00	10	1.00	10	11.00	66.00	0.0016	0.0031APPROX
5.00	10	1.00	10	7.00	62.00	0.0006	0.0011APPROX
6.00	10	1.00	10	1.00	56.00	0.0001	0.0002APPROX
3.00	10	2.00	10	28.50	83.50	0.0517	0.1033APPROX
4.00	10	2.00	10	29.00	84.00	0.0560	0.1120APPROX
5.00	10	2.00	10	25.00	80.00	0.0292	0.0585APPROX
6.00	10	2.00	10	25.00	80.00	0.0292	0.0585APPROX
4.00	10	3.00	10	47.50	102.50	0.4249	0.8498APPROX
5.00	10	3.00	10	34.50	89.50	0.1204	0.2408APPROX
6.00	10	3.00	10	23.00	78.00	0.0205	0.0409APPROX
5.00	10	4.00	10	42.50	97.50	0.2849	0.5699APPROX
6.00	10	4.00	10	19.50	74.50	0.0104	0.0207APPROX
6.00	10	5.00	10	30.00	85.00	0.0650	0.1300APPROX

MANN-WHITNEY U FOR GROUP VARIABLE = GRADE SCORE VARIABLE = S2B

WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	10	0.0	10	27.50	82.50	0.0322	0.0643
2.00	10	0.0	10	21.00	76.00	0.0092	0.0183
3.00	10	0.0	10	3.50	58.50	0.0000	0.0001
4.00	10	0.0	10	3.50	58.50	0.0000	0.0001
5.00	10	0.0	10	0.0	55.00	0.0000	0.0000
6.00	10	0.0	10	0.0	55.00	0.0000	0.0000
2.00	10	1.00	10	42.00	97.00	0.2670	0.5340
3.00	10	1.00	10	29.50	84.50	0.0572	0.1144
4.00	10	1.00	10	24.50	79.50	0.0209	0.0418
5.00	10	1.00	10	20.00	75.00	0.0054	0.0108
6.00	10	1.00	10	20.00	75.00	0.0054	0.0108
3.00	10	2.00	10	41.50	96.50	0.2399	0.4799
4.00	10	2.00	10	34.53	89.50	0.0975	0.1950
5.00	10	2.00	10	30.00	85.00	0.0433	0.0867
6.00	10	2.00	10	30.00	85.00	0.0433	0.0867
4.00	10	3.00	10	41.00	96.00	0.2910	0.5820
5.00	10	3.00	10	35.00	90.00	0.1053	0.2105
6.00	10	3.00	10	35.00	90.00	0.1053	0.2105
5.00	10	4.00	10	45.00	100.00	0.5000	1.0000
6.00	10	4.00	10	45.00	100.00	0.5000	1.0000
5.00	10	6.00	10	50.00	105.00	0.0	0.0

MANN-WHITNEY U FOR GROUP VARIABLE = GRADE SCORE VARIABLE = S1B

WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
0.0	10	1.00	10	49.50	104.50	0.5000	1.0000
0.0	10	2.00	10	49.50	104.50	0.5000	1.0000
3.00	10	0.0	10	45.00	100.00	0.5000	1.0000
4.00	10	0.0	10	45.00	100.00	0.5000	1.0000
5.00	10	0.0	10	45.00	100.00	0.5000	1.0000
6.00	10	0.0	10	45.00	100.00	0.5000	1.0000
1.00	10	2.00	10	50.00	105.00	0.7632	1.0000
3.00	10	1.00	10	45.00	100.00	0.5000	1.0000
4.00	10	1.00	10	45.00	100.00	0.5000	1.0000
5.00	10	1.00	10	45.00	100.00	0.5000	1.0000
6.00	10	1.00	10	45.00	100.00	0.5000	1.0000
3.00	10	2.00	10	45.00	100.00	0.5000	1.0000
4.00	10	2.00	10	45.00	100.00	0.5000	1.0000
5.00	10	2.00	10	45.00	100.00	0.5000	1.0000
6.00	10	2.00	10	45.00	100.00	0.5000	1.0000
3.00	10	4.00	10	50.00	105.00	0.0	0.0
3.00	10	5.00	10	50.00	105.00	0.0	0.0
3.00	10	6.00	10	50.00	105.00	0.0	0.0
4.00	10	5.00	10	50.00	105.00	0.0	0.0
4.00	10	6.00	10	50.00	105.00	0.0	0.0
5.00	10	6.00	10	50.00	105.00	0.0	0.0

MANN-WHITNEY U FOR GROUP VARIABLE = SEX
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	32	2.00	38	553.50	1294.50	0.2600	0.5200APPROX

SCORE VARIABLE = S1A

MANN-WHITNEY U FOR GROUP VARIABLE = SEX
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
2.00	38	1.00	32	506.00	1034.00	0.1143	0.2287APPROX

SCORE VARIABLE = S2A

MANN-WHITNEY U FOR GROUP VARIABLE = SEX
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
2.00	38	1.00	32	575.00	1103.00	0.3485	0.6971APPROX

SCORE VARIABLE = TOTAL

MANN-WHITNEY U FOR GROUP VARIABLE = SEX
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
1.00	32	2.00	38	490.50	1231.50	0.0500	0.1001APPROX

SCORE VARIABLE = S2B

MANN-WHITNEY U FOR GROUP VARIABLE = SEX
WITH PROBABILITY OF THIS OR SMALLER U

HIGHER GROUP	N1	LOWER GROUP	N2	U	RANK SUM	PROB (1-TAIL)	(2-TAIL)
2.00	38	1.00	32	586.50	1114.50	0.2350	0.4701APPROX

SCORE VARIABLE = S1B

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = GRADE SCORE VARIABLE = S1A

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
0.0	10.	47.500	4.000	1.00	10.	47.900	3.105	-0.2370	0.8153	1.6597	0.23
0.0	10.	47.500	4.000	2.00	10.	48.750	4.468	-0.6253	0.5396	0.8015	0.63
0.0	10.	47.500	4.000	3.00	10.	52.250	3.516	-2.6757	0.0154	1.2942	0.35
0.0	10.	47.500	4.000	4.00	10.	51.250	1.721	-2.5835	0.0187	5.4008	0.01
0.0	10.	47.500	4.000	5.00	10.	52.000	4.231	-2.3186	0.0324	0.8939	0.57
0.0	10.	47.500	4.000	6.00	10.	52.950	2.030	-3.6448	0.0019	3.8810	0.03
1.00	10.	47.900	3.105	2.00	10.	48.750	4.468	-0.4687	0.6449	0.4829	0.85
1.00	10.	47.900	3.105	3.00	10.	52.250	3.516	-2.7821	0.0123	0.7798	0.64
1.00	10.	47.900	3.105	4.00	10.	51.250	1.721	-2.8310	0.0111	3.2541	0.05
1.00	10.	47.900	3.105	5.00	10.	52.000	4.231	-2.3438	0.0308	0.5386	0.81
1.00	10.	47.900	3.105	6.00	10.	52.950	2.030	-4.0837	0.0007	2.3384	0.11
2.00	10.	48.750	4.468	3.00	10.	52.250	3.516	-1.8468	0.0813	1.6148	0.24
2.00	10.	48.750	4.468	4.00	10.	51.250	1.721	-1.5664	0.1347	6.7384	0.00
2.00	10.	48.750	4.468	5.00	10.	52.000	4.231	-1.5845	0.1305	1.1152	0.44
2.00	10.	48.750	4.468	6.00	10.	52.950	2.030	-2.5674	0.0194	4.8421	0.01
3.00	10.	52.250	3.516	4.00	10.	51.250	1.721	0.7663	0.4534	4.1730	0.02
3.00	10.	52.250	3.516	5.00	10.	52.000	4.231	0.1363	0.8931	0.6906	0.70
3.00	10.	52.250	3.516	6.00	10.	52.950	2.030	-0.5172	0.6113	2.9987	0.06
4.00	10.	51.250	1.721	5.00	10.	52.000	4.231	-0.4926	0.6282	0.1655	0.99
4.00	10.	51.250	1.721	6.00	10.	52.950	2.030	-1.9160	0.0714	0.7186	0.68
5.00	10.	52.000	4.231	6.00	10.	52.950	2.030	-0.6073	0.5512	4.3419	0.02

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = GRADE SCORE VARIABLE = S2A

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
0.0	10.	47.100	3.555	1.00	10.	48.450	2.936	-0.8783	0.3914	1.4659	0.29
0.0	10.	47.100	3.555	2.00	10.	47.550	4.917	-0.2225	0.8264	0.5229	0.83
0.0	10.	47.100	3.555	3.00	10.	49.850	3.627	-1.6244	0.1217	0.9610	0.52
0.0	10.	47.100	3.555	4.00	10.	51.050	3.503	-2.3741	0.0289	1.0300	0.48
0.0	10.	47.100	3.555	5.00	10.	52.200	3.132	-3.2291	0.0047	1.2865	0.36
0.0	10.	47.100	3.555	6.00	10.	52.750	3.084	-3.6013	0.0020	1.3288	0.34
1.00	10.	48.450	2.936	2.00	10.	47.550	4.917	0.4715	0.6430	0.3567	0.93
1.00	10.	48.450	2.936	3.00	10.	49.850	3.627	-0.9000	0.3800	0.6556	0.73
1.00	10.	48.450	2.936	4.00	10.	51.050	3.503	-1.7064	0.1051	0.7026	0.70
1.00	10.	48.450	2.936	5.00	10.	52.200	3.132	-2.6203	0.0173	0.8790	0.57
1.00	10.	48.450	2.936	6.00	10.	52.750	3.084	-3.0292	0.0072	0.9065	0.56
2.00	10.	47.550	4.917	3.00	10.	49.850	3.627	-1.1294	0.2736	1.8378	0.19
2.00	10.	47.550	4.917	4.00	10.	51.050	3.503	-1.7393	0.0991	1.9696	0.16
2.00	10.	47.550	4.917	5.00	10.	52.200	3.132	-2.3930	0.0278	2.4640	0.10
2.00	10.	47.550	4.917	6.00	10.	52.750	3.084	-2.6879	0.0150	2.5411	0.09
3.00	10.	49.850	3.627	4.00	10.	51.050	3.503	-0.7139	0.4844	1.0717	0.46
3.00	10.	49.850	3.627	5.00	10.	52.200	3.132	-1.4712	0.1585	1.3427	0.33
3.00	10.	49.850	3.627	6.00	10.	52.750	3.084	-1.8274	0.0843	1.3927	0.32
4.00	10.	51.050	3.503	5.00	10.	52.200	3.132	-0.7342	0.4723	1.2510	0.37
4.00	10.	51.050	3.503	6.00	10.	52.750	3.084	-1.0927	0.2889	1.2902	0.36
5.00	10.	52.200	3.132	6.00	10.	52.750	3.084	-0.3754	0.7118	1.0313	0.48

157

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = GRADE SCORE VARIABLE = TOTAL

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
0.0	10.	172.900	5.024	1.00	10.	175.400	4.711	-1.0889	0.2905	1.1374	0.43
0.0	10.	172.900	5.024	2.00	10.	175.500	9.418	-0.7367	0.4744	0.2846	0.96
0.0	10.	172.900	5.024	3.00	10.	181.900	3.872	-4.2565	0.0005	1.6836	0.22
0.0	10.	172.900	5.024	4.00	10.	182.200	2.952	-4.7879	0.0001	2.8974	0.06
0.0	10.	172.900	5.024	5.00	10.	184.200	5.533	-4.5359	0.0003	0.8246	0.61
0.0	10.	172.900	5.024	6.00	10.	185.700	2.704	-6.7300	0.0000	3.4521	0.04
1.00	10.	175.400	4.711	2.00	10.	175.500	9.418	-0.0285	0.9776	0.2502	0.97
1.00	10.	175.400	4.711	3.00	10.	181.900	3.872	-3.1977	0.0050	1.4802	0.28
1.00	10.	175.400	4.711	4.00	10.	182.200	2.952	-3.6695	0.0018	2.5473	0.07
1.00	10.	175.400	4.711	5.00	10.	184.200	5.533	-3.6330	0.0019	0.7250	0.69
1.00	10.	175.400	4.711	6.00	10.	185.700	2.704	-5.6885	0.0000	3.0350	0.06
2.00	10.	175.500	9.418	3.00	10.	181.900	3.872	-1.8855	0.0756	5.9158	0.01
2.00	10.	175.500	9.418	4.00	10.	182.200	2.952	-2.0365	0.0567	10.1838	0.00
2.00	10.	175.500	9.418	5.00	10.	184.200	5.533	-2.3894	0.0280	2.8975	0.06
2.00	10.	175.500	9.418	6.00	10.	185.700	2.704	-3.1229	0.0059	12.1299	0.00
3.00	10.	181.900	3.872	4.00	10.	182.200	2.952	-0.1848	0.8554	1.7209	0.22
3.00	10.	181.900	3.872	5.00	10.	184.200	5.533	-0.3217	0.3204	0.4878	0.85
3.00	10.	181.900	3.872	6.00	10.	185.700	2.704	-2.4137	0.0267	2.0504	0.15

150

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
4.00	10.	182.200	2.952	5.00	10.	184.200	5.533	-0.9568	0.3513	0.2846	0.96
4.00	10.	182.200	2.952	6.00	10.	185.700	2.704	-2.6230	0.0172	1.1915	0.40
5.00	10.	184.200	5.533	6.00	10.	185.700	2.704	-0.7307	0.4744	4.1863	0.02

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = GRADE SCORE VARIABLE = S2B

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
0.0	10.	38.350	1.484	1.00	10.	39.250	0.750	-1.6237	0.1218	3.9160	0.03
0.0	10.	38.350	1.484	2.00	10.	39.400	0.889	-1.8208	0.0853	2.7874	0.07
0.0	10.	38.350	1.484	3.00	10.	39.800	0.332	-2.8603	0.0104	19.9964	0.00
0.0	10.	38.350	1.484	4.00	10.	39.900	0.300	-3.0708	0.0066	24.4113	0.00
0.0	10.	38.350	1.484	5.00	10.	40.000	0.0	-3.3352	0.0037	0.0	1.00
0.0	10.	38.350	1.484	6.00	10.	40.000	0.0	-3.3352	0.0037	0.0	1.00
1.00	10.	39.250	0.750	2.00	10.	39.400	0.889	-0.3869	0.7034	0.7118	0.69
1.00	10.	39.250	0.750	3.00	10.	39.800	0.332	-2.0118	0.0595	5.1064	0.01
1.00	10.	39.250	0.750	4.00	10.	39.900	0.300	-2.4136	0.0267	6.2338	0.01
1.00	10.	39.250	0.750	5.00	10.	40.000	0.0	-3.0000	0.0077	0.0	1.00
1.00	10.	39.250	0.750	6.00	10.	40.000	0.0	-3.0000	0.0077	0.0	1.00
2.00	10.	39.400	0.889	3.00	10.	39.800	0.332	-1.2646	0.2221	7.1738	0.00
2.00	10.	39.400	0.889	4.00	10.	39.900	0.300	-1.5986	0.1273	8.7576	0.00
2.00	10.	39.400	0.889	5.00	10.	40.000	0.0	-2.0249	0.0580	0.0	1.00
2.00	10.	39.400	0.889	6.00	10.	40.000	0.0	-2.0249	0.0580	0.0	1.00
3.00	10.	39.800	0.332	4.00	10.	39.900	0.300	-0.6702	0.5112	1.2208	0.39
3.00	10.	39.800	0.332	5.00	10.	40.000	0.0	-1.8079	0.0874	0.0	1.00
3.00	10.	39.800	0.332	6.00	10.	40.000	0.0	-1.8079	0.0874	0.0	1.00
4.00	10.	39.900	0.300	5.00	10.	40.000	0.0	-0.9988	0.3311	0.0	1.00
4.00	10.	39.900	0.300	6.00	10.	40.000	0.0	-0.9988	0.3311	0.0	1.00
5.00	10.	40.000	0.0	6.00	10.	40.000	0.0	0.0	1.0000	0.0	1.00

158

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = GRADE SCORE VARIABLE = S1B

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
0.0	10.	39.950	0.151	1.00	10.	39.800	0.600	0.7274	0.4764	0.0629	1.00
0.0	10.	39.950	0.151	2.00	10.	39.800	0.600	0.7274	0.4764	0.0629	1.00
0.0	10.	39.950	0.151	3.00	10.	40.000	0.0	-0.9966	0.3322	0.0	1.00
0.0	10.	39.950	0.151	4.00	10.	40.000	0.0	-0.9966	0.3322	0.0	1.00
0.0	10.	39.950	0.151	5.00	10.	40.000	0.0	-0.9966	0.3322	0.0	1.00
0.0	10.	39.950	0.151	6.00	10.	40.000	0.0	-0.9966	0.3322	0.0	1.00
1.00	10.	39.800	0.600	2.00	10.	39.800	0.600	0.0	1.0000	1.0000	0.50
1.00	10.	39.800	0.600	3.00	10.	40.000	0.0	-0.9998	0.3306	0.0	1.00
1.00	10.	39.800	0.600	4.00	10.	40.000	0.0	-0.9998	0.3306	0.0	1.00
1.00	10.	39.800	0.600	5.00	10.	40.000	0.0	-0.9998	0.3306	0.0	1.00
1.00	10.	39.800	0.600	6.00	10.	40.000	0.0	-0.9998	0.3306	0.0	1.00
2.00	10.	39.800	0.600	3.00	10.	40.000	0.0	-0.9998	0.3306	0.0	1.00
2.00	10.	39.800	0.600	4.00	10.	40.000	0.0	-0.9998	0.3306	0.0	1.00

151

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
2.00	10.	39.800	0.600	5.00	10.	40.000	0.0	-0.9998	0.3306	0.0	1.00
2.00	10.	39.800	0.600	6.00	10.	40.000	0.0	-0.9998	0.3306	0.0	1.00
3.00	10.	40.000	0.0	4.00	10.	40.000	0.0	0.0	1.0000	0.0	1.00
3.00	10.	40.000	0.0	5.00	10.	40.000	0.0	0.0	1.0000	0.0	1.00
3.00	10.	40.000	0.0	6.00	10.	40.000	0.0	0.0	1.0000	0.0	1.00
4.00	10.	40.000	0.0	5.00	10.	40.000	0.0	0.0	1.0000	0.0	1.00
4.00	10.	40.000	0.0	6.00	10.	40.000	0.0	0.0	1.0000	0.0	1.00
5.00	10.	40.000	0.0	6.00	10.	40.000	0.0	0.0	1.0000	0.0	1.00

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = SEX SCORE VARIABLE = S1A

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
2.00	38.	50.039	4.032	1.00	32.	50.766	3.986	-0.7437	0.4596	1.0230	0.48

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = SEX SCORE VARIABLE = S2A

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
2.00	38.	50.355	4.996	1.00	32.	49.250	4.127	1.1047	0.2732	0.9850	0.52

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = SEX SCORE VARIABLE = TOTAL

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
2.00	38.	179.684	6.919	1.00	32.	179.688	7.194	-0.0019	0.9985	0.9250	0.59

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = SEX SCORE VARIABLE = S2B

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
2.00	38.	39.342	1.142	1.00	32.	39.750	0.451	-1.8725	0.0654	6.4200	0.00

FISHER-STUDENT T-TEST AND SNEDECOR-FISHER F-TEST
WITH PROBABILITIES OF LARGER T AND F

GROUP VARIABLE = SEX SCORE VARIABLE = SIB

GROUP 1	N1	MEAN	STD DEV	GROUP 2	N2	MEAN	STD DEV	T	PROB	F	PROB
2.00	38.	39.947	0.320	1.00	32.	39.922	0.356	0.3104	0.7572	0.8081	0.73

BIBLIOGRAPHY

Agard, F. B., and H. B. Dunkel, An Investigation of Second Language Teaching, Boston: Ginn & Co., 1948.

"A Million Random Digits With 100,000 Normal Deviates," Glencoe, Illinois: The Free Press, 1955.

Arendt, Jermaine D., "New Scheduling Patterns and the Foreign Language Teacher," ERIC Focus Reports on the Teaching of Foreign Languages, No. 18, New York: The Modern Language Association of America, 1970.

Asher, James J., and Ramiro Garcia, "The Optional Age to Learn a Foreign Language," Modern Language Journal, 53:334-341, May, 1969.

Ausubel, David P., "Adults Versus Children in Second-Language Learning: Psychological Consideration," reprinted in: Mildred R. Donoghue (ed.), Foreign Languages and the Schools: A Book of Readings, (pp. 100-106) Dubuque, Iowa: Wm. C. Brown Company Publishers, 1967; Modern Language Journal, 48:420-424, November, 1964.

Boehm, Leonore, "Age and Foreign Language Training," Modern Language Journal, 43:32-33, January, 1959.

Bourque, Edward H. (ed.), The FLES Student: A Study, Philadelphia: Chilton Books, Educational Division, 1968.

Brega, Evelyn, and John M. Newell, "Comparison of Performance by FLES Program Students and Regular French III Students on Modern Language Association Tests," French Review, 39:433-438, December, 1965.

Brega, Evelyn, and John M. Newell, "High-School Performance of FLES and Non-FLES Students," Modern Language Journal, 51:408-411, November, 1967.

Brooks, Nelson, Language and Language Learning: Theory and Practice (2nd ed.), New York, Harcourt, Brace and World, Inc., 1964.

Carroll, John B., "Foreign Languages for Children: What Research Says," National Elementary Principal, 39: 12-15, May, 1960.

Carroll, John B., "Foreign Language Proficiency Levels Attained by Language Majors near Graduation from College," Foreign Language Annals, 1:131-151, December, 1967.

Carroll, John B., "Research on Teaching Foreign Languages," in: N. L. Gage (ed.), Handbook of Research on Teaching, (pp. 1060-1100) Chicago: Rand McNally & Co., 1963.

Carroll, John B., "Wanted: A Research Basis for Educational Policy on Foreign Language Teaching," Harvard Educational Review, 30:128-140, Spring, 1960.

Carroll, John B., and Stanley M. Sapon, "Modern Language Aptitude Test: Manual," New York: The Psychological Corporation, 1959.

Cronbach, L.J., "The Role of the University in Improving Education," Phi Delta Kappan, 47:539-545, June, 1966.

Denemark, George W., and Wesley J. Matson, "Why Teach Children a Foreign Language?" National Elementary Principal, 39:6-11, May, 1960.

Donoghue, Mildred R., "Foreign Languages in the Elementary School: Effects and Instructional Arrangements According to Research," ERIC Focus Reports on the Teaching of Foreign Languages, No. 3, New York: The Modern Language Association of America, 1969.

Dunkel, Harold B., and Roger A. Pillet, French in the Elementary School: Five Years' Experience, Chicago: The University of Chicago Press, 1962.

"Foreign Language Bulletins," No. 1, New York: Modern Language Association of America, April, 1953.

"Foreign Languages in the Elementary School: A Statement of Policy," New York: Modern Language Association, 1961.

Gleason, H.A., Jr., An Introduction to Descriptive Linguistics (Revised Edition), New York: Holt, Rinehart and Winston, 1967.

Goodman, Allan Cooper, "Imitation of Intonation Patterns," Unpublished Ph.D. Dissertation, University of Michigan, Ann Arbor, 1952.

Gourevitch, Vivian. Statistical Methods: A Problem-Solving Approach. Boston: Allyn and Bacon, Inc., 1965.

Grinder, R.E., A Otomo and W. Toyota, "Comparisons Between Second, Third and Fourth-Grade Children in the Audio-Lingual Learning of Japanese as a Second Language," Journal of Educational Research, 56:191-197, December, 1962.

Grittner, Frank, M., Teaching Foreign Languages, New York: Harper & Row, 1969.

Haas, Werner, "Is the Four-Skills Approach Obsolete?" Die Unterrichtspraxis, 3:61-65, Fall, 1970.

Hildreth, Gertrude, "Learning a Second Language in the Elementary Grades and High School," Modern Language Journal, 43:136-142, March, 1959.

JeKenta, Albert W., and Percy Fearing, "Current Trends in Curriculum: Elementary and Secondary Schools," in: Emma Birkmaier (ed.), Britannica Review of Foreign Language Education, (Chap. 6, pp. 141-178), Chicago: Encyclopaedia Britannica, Inc. 1968.

Justman, Joseph, and Martin L. Nass, "The High School Achievement of Pupils Who Were and Were not Introduced to a Foreign Language in Elementary School," Modern Language Journal, 40:120-123, March, 1956.

Kirch, Max S., "At What Age Elem ntary School Language Teaching," Modern Language Journal, 40 399-400, November, 1956.

Kirch, Max S., "Foreign Languages in the Elementary Schools: First Grade German," Modern Language Journal, 39:144-145, March, 1955.

Lado, Robert, Language Testing, New York: McGraw-Hill Book Company, 1964.

Lange, Dale L. (ed.), Britannica Review of Foreign Language Education (Vol. 2), Chicago: Encyclopaedia Britannica, Inc., 1970.

Larew, Leonor A. "A Study of Spanish Articulation in the Elementary School: A Pilot Study," Unpublished Ph.D. Dissertation, University of Missouri, Columbia, 1960.

Larew, Leonor A., "The Optimum Age for Beginning a Foreign Language," Modern Language Journal, 45:203-206, May, 1961.

Mase, Darrell J., "Disorders of Articulation," in N. M. Levin (ed.), Voice and Speech Disorders: Medical Aspects, Springfield, Illinois: Charles C. Thomas, Publisher, 1962.

Michel, Joseph (ed.), Foreign Language Teaching: An Anthology, New York: The Macmillan Company, 1967.

Miel, Alice "Does Foreign Language Belong in Elementary School?" Teachers College Record, 56:139-148, December, 1954.

"MLA Cooperative Foreign Language Tests: German," Educational Testing Service, Cooperative Test Division, Princeton, New Jersey: 1965.

Penfield, Wilder, "A Consideration of the Neuro-Physiological Mechanisms of Speech and Some Educational Consequences," Bulletin of the American Academy of Arts and Sciences, 82:199-214, Boston: April 5, 1953.

Penfield, Wilder, "The Uncommitted Cortex," The Atlantic Monthly, 213:77-81, July, 1964.

Penfield, Wilder and Lamar Roberts, Speech and Brain-Mechanisms, Princeton, N.J.: Princeton University Press, 1959.

Pimsleur, Paul, "Pimsleur German Proficiency Tests," New York: Harcourt, Brace & World, Inc., 1967.

Pimsleur, Paul, et al., Under-Achievement in Foreign Language Learning, New York: The Modern Language Association of America, 1966.

Rivers, Wilga M., Teaching Foreign Language Skills, Chicago: The University of Chicago Press, 1968.

Roe, Vivian, and R. Milisen, "The Effect of Maturation Upon Defective Articulation in Elementary Grades," Journal of Speech Disorders, 7:37-45, March, 1942.

Sayler, Helen K., "The Effect of Maturation Upon Defective Articulation in Grades Seven through Twelve," Journal of Speech and Hearing Disorders, 14:202-207, September, 1949.

Smith, Philip D. Jr., A Comparison of the Cognitive and Audio-lingual Approaches to Foreign Language Instruction, Philadelphia: The Center for Curriculum Development, Inc., 1970.

Stern, H.H., "FLES: Achievements and Problems," in: Albert Valdman (ed.), Trends in Language Teaching, (pp. 253-283), New York: McGraw-Hill Book Co., 1966.

"Upper Arlington Schools," a brochure available from: The Superintendent's Office, 1950 North Mallway, Columbus, Ohio, 43221.

Van Dalen, Deobold B. with two chapter by Meyer, William J., Understanding Educational Research, New York: McGraw-Hill Book Company, 1966.

Vocolo, Joseph M. "The Effect of Foreign Language Study in the Elementary School Upon Achievement in the Same Foreign Language in the High School," Modern Language Journal, 51:463-469, December, 1967.

Walsh, Donald D., "What's What: A List of Useful Terms for the Teacher of Modern Languages," (2nd ed.) New York: The Modern Language Associatior of America, 1964.

Wängler, Hans-Heinrich, Instruction in German Pronunciation, St. Paul, Minnesota: EMC Corporation, 1963.

Wängler, Hans-Heinrich, Patterns in German Stress and Intonation, St. Paul, Minnesota: EMC Corporation, 1966.

Winitz, Harris and Martha Lawrence, "Children's Articulation and Sound Learning Ability," Journal of Speech and Hearing Research, 4:259-268, Sept. 1961.